

UNIVERSIDAD MAYOR DE SAN ANDRÉS
Facultad de Ciencias Puras y Naturales
Carrera de Estadística



Tesis de Grado

**APLICACIÓN DEL MODELO
DE MARKOV DIFUSO**

TUTOR: LIC. YURI MIRANDA GONZALES
POSTULANTE: RONAL EDWIN CONDORI HUANCA

La Paz – Bolivia
2013

DEDICATORIA

*A la memoria de mi madre (†), que constantemente
supo guiarme en todo momento de su vida.*

AGRADECIMIENTOS

Agradecer al Lic. Yuri Miranda por su guía en el desarrollo del presente trabajo, al M.Sc. Franz Cuevas por haber cubierto diversas incertidumbres en el campo de investigación de la teoría difusa, y un reconocimiento en particular al Lic. Fernando Rivero, M.Sc. Franklin Garcia, y Ph.D. Manighe Roosta, por el apoyo y ánimo brindado desde un inicio.

*Del hablador he aprendido a callar; del intolerante,
a ser indulgente, y del malévolo a tratar a los demás con amabilidad.
Y por curioso que parezca, no siento ninguna gratitud hacia estos maestros.*

Khalil Gibran (1883-1991).

RESUMEN

La teoría difusa o *fuzzy*, al igual que la teoría de probabilidad posee orígenes matemáticos con un enfoque de estudio a la incertidumbre, compartiendo herramientas, y compitiendo fuertemente en áreas de aplicación. Desde su inicio (Zadeh, 1965) la teoría *fuzzy* ha despertado fuertes reprochas, y de la misma manera las sus aplicaciones sobre esta, en particular nos referimos a las series de tiempo difusas (Song & Chissom, 1993) siendo una alternativa para el manejo de datos temporales.

En el capítulo primero se exponen los orígenes de la Teoría Difusa junto a los conjuntos; las Series de Tiempo Difusas en sus inicios; y finalmente se establecen los objetivos a ser alcanzados para la presente tesis.

El segundo capítulo se abarcan los Conjuntos, Números Difusos y sus propiedades; se definen formalmente las Variables Aleatorias Difusas, se identifican Modelos de Series de Tiempo Difusas, medias generalizadas, y los Métodos de Estimación de Cadenas de Markov incorporando finalmente el uso de la Distribución Dirichlet.

El capítulo final, presenta la serie de pasos metodológicos, en la modelación de datos provenientes para las denominadas AFP'S, su desarrollo comparativo con el modelo ARIMA, junto a las conclusiones y recomendaciones pertinentes.

ÍNDICE

<i>Dedicatoria</i>	<i>II</i>
<i>Agradecimientos</i>	<i>III</i>
<i>Resumen</i>	<i>V</i>
<i>Índice</i>	<i>VI</i>
<i>Índice de ilustraciones</i>	<i>IX</i>
<i>Índice de tablas</i>	<i>X</i>

Capítulo I: Nociones Generales

1.1. Introducción	1
1.2. Antecedentes	2
1.2.1. Origen de los Conjuntos Difusos	2
1.2.2. Conjuntos Difusos.....	3
1.2.3. Variables Lingüísticas	8
1.2.4. Serie de Tiempo Difusa.....	8
1.3. Planteamiento del Problema	10
1.4. Hipótesis	10
1.5. Objetivos	10
1.5.1. Objetivo General.....	10
1.5.2. Objetivos Específicos	10
1.6. Justificación.....	11
1.7. Alcances y Limitaciones	11

Capítulo II: Marco Teórico

2.1. Conjuntos Difusos	12
2.1.1. Funciones de Pertenencia	14
2.1.2. Operadores entre Conjuntos Difusos.....	15
Relaciones Difusas.....	18
2.1.3. Números Difusos	19
Operaciones Aritméticas entre Números Difusos	19
2.1.4. Variables Aleatorias Difusas	21
2.1.5. Métodos de Defuzzificación	25
2.1.6. Conjuntos Difusos sobre Variables Lingüísticas.....	26
2.1.7. Series de Tiempo Difusas	27

2.2. Modelos de Series de Tiempo Difusas	30
2.2.1. Etapa 1ra: Creación de Clusters & Relaciones Difusas	30
2.2.2. Etapa 2da: Determinación de Relaciones	31
Modelo de Chen (1996).....	32
Modelo de Yu's (2005)	33
Modelo de Lee's (2008)	34
2.2.3. Etapa 3ra Calculo de Pronósticos.....	34
2.2.4. Críticas y Avances en las STD.....	34
2.2.5. Medidas de Bondad de Ajuste	35
2.3. Medias Generalizadas	36
2.3.1. La F-media Generalizada	36
Propiedades de la F-media Generalizada.....	37
2.3.2. Medias Ponderadas	38
2.3.3. Medias Ordenadas Ponderadas (OWA).....	39
2.3.4. Aplicación de la F-media en las Series de Tiempo Difusas	39
2.4. Cadenas de Markov	40
2.4.1. Matriz de Transición	42
2.4.2. Método de Estimación por Máxima Verosimilitud	43
Método de Máxima Verosimilitud: Propiedades.....	44
Método de Máxima Verosimilitud: Generalización	45
Prueba de Homogeneidad para Cadenas de Markov	46
2.4.3. Método de Estimación Bayesiano.....	46
Distribución Beta	47
Elección de la Distribución Priori	49
Elección del Estimador Puntual	50
Distribución Dirichlet.....	50
Priori Multivariante de Jeffreys	52
2.5. Aplicación de Cadenas de Markov	55

Capítulo III: Marco Practico

3.1. Preliminares.....	56
3.1.1. Reforma al sistema de reparto.....	56
3.1.2. El nuevo sistema de reparto.....	57
3.1.3. Las administradoras de fondos de pensiones	58
3.2. Aplicación del modelo de markov difuso.....	59
3.2.1. Construcción de los clústers	60
3.2.2. Estimación de la matriz de markov	61
3.2.2. Números y variables aleatorias difusas	63
3.2.3. Calculo de pronósticos	65
3.2.4. Medidas de bondad de ajuste	68
3.3. Metodología box & jenkins: modelo arima	68
3.4. Comparación entre modelo de markov difuso y arima.....	69
3.5. Conclusiones	70
3.5.1. Respecto al modelo de markov difuso.....	70
3.5.2. Respecto a los resultados.....	71
3.6. Recomendaciones	71

Anexos

A.1. Sigma Algebra & Teoría de la Probabilidad.....	72
A.1.1. Nociones Sobre Variables Aleatorias Difusas.....	74
A.2. Números Difusos	76
A.2.1. Operadores Aritméticos.....	76
A.2.2. Distancia entre Números Difusos.....	79
A.3. Distribuciones y Propiedades	80
A.3.1. Distribución Dirichlet.....	80
A.3.2. Distribución Trapezoidal	83
A.4. Metodología BOX & JENKINS: Modelo ARIMA	87
A.5. Do-file Stata 12.0.....	88
A.6. Datos y Resultados Obtenidos	91
A.6.1. Para el Modelo de Markov.....	92
A.6.2. Para el Modelo ARIMA	93
BIBLIOGRAFÍA.....	97

ÍNDICE DE ILUSTRACIONES

ILUSTRACIÓN 1. REPRESENTACIÓN DE LA FUNCIÓN DE PERTENENCIA: $1 - e^{-x/100}$	5
ILUSTRACIÓN 2. DIFERENTES OPERACIONES: $A, B, A \cup B$ Y $A \cap B$	6
ILUSTRACIÓN 3. REPRESENTACIÓN GRÁFICA DE ALFA-CORTES	7
ILUSTRACIÓN 4. REPRESENTACIÓN DE DIFERENTES FUNCIONES DE PERTENENCIA.....	7
ILUSTRACIÓN 5. LOFTY ZADEH EL DENOMINADO PADRE DE LA TEORÍA FUZZY	12
ILUSTRACIÓN 6. NÚMERO DIFUSO TRAPEZOIDAL EN TÉRMINOS DE CUARTETAS	21
ILUSTRACIÓN 7. REPRESENTACIÓN DE UNA VARIABLE DIFUSA	24
ILUSTRACIÓN 8. RELACIÓN DIFUSA DE 1ER ORDEN	29
ILUSTRACIÓN 9. RELACIÓN DIFUSA DE ORDEN K	29
ILUSTRACIÓN 10. RELACIÓN DIFUSA DE GRUPO	30
ILUSTRACIÓN 11. PRIMER Y SEGUNDO CASO EN LA REGLA DE CHEN	32
ILUSTRACIÓN 12. TERCER CASO EN LA REGLA DE CHEN	33
ILUSTRACIÓN 13. ANTHONY B. ATKINSON AUTOR DE: “ <i>ON THE MEASUREMENT OF INEQUALITY</i> ”	36
ILUSTRACIÓN 14. ANDREY MARKOV (RUSIA, 1856-1922).....	41
ILUSTRACIÓN 15. DENSIDADES BETA PARA DISTINTOS VALORES DE α Y β	48
ILUSTRACIÓN 16. CURVAS DE NIVEL PARA DISTINTAS DENSIDADES DIRICHLET	51
ILUSTRACIÓN 17. DENSIDAD PRIORI POCO INFORMATIVA $Dir(\alpha_1 = \alpha_2 = \alpha_3 = 1/2)$	55
ILUSTRACIÓN 18. SERIE DE DATOS 06-1997 A 12-2011	59
ILUSTRACIÓN 19. SERIE DE DATOS TRANSFORMADOS W_T	59
ILUSTRACIÓN 20. CLUSTERS CONSTRUIDOS ($k=14$).....	60
ILUSTRACIÓN 21. REPRESENTACIÓN DE LOS CLUSTERS CONFORMADOS (A3, A4 & A5)	64
ILUSTRACIÓN 22. VALORES PRONOSTICADOS MEDIANTE NÚMEROS DIFUSOS	67
ILUSTRACIÓN 23. SERIE PRONOSTICADA DEFUZZIFICADA Y SERIE REAL	67
ILUSTRACIÓN 24. SERIE ESTIMADA ARIMA (0;1;1) SARIMA(1,1,0) Y SERIE REAL	69
ILUSTRACIÓN 25. EVOLUCIÓN DE ECM PARA DIFERENTES NÚMEROS DE CLUSTERS	70

ÍNDICE DE TABLAS

TABLA 1. CLASIFICACION DE MEDIAS ORDINARIAS	39
TABLA 2. CLASIFICACION DE PROCESOS DE MARKOV	41
TABLA 3. DESCRIPCIÓN DE LOS CLUSTERS CONFORMADOS	60
TABLA 4. RELACIONES DIFUSAS AGRUPADAS	61
TABLA 5. APROXIMACIÓN DE NÚMEROS DIFUSOS TRAPEZOIDALES	65
TABLA 6. VALORES ESPERADOS TRAPEZOIDALES	66
TABLA 7. COMPARACION DE ECM ENTRE MODELOS.....	69
TABLA 8. SERIE DE DATOS MANEJADOS	91

Capítulo I

Nociones Generales

1.1. Introducción

La mayoría de los modelos estadísticos con los que se trabaja, se plantean y desarrollan bajo un dominio de precisión sobre las variables a ser medidas, estas mismas intervienen en todo el proceso de estimación y el cálculo de los parámetros que definen concretamente a cada modelo.

Bajo esta realidad se obvian las imprecisiones que presentan la mayoría de los datos a ser analizados, debido a esto si el objetivo es conseguir resultados más representativos, es necesario la incorporación de modelos en donde tanto los datos y como los parámetros sean *insuficientes, imperfectos o difusos*¹, los mismos datos serán originados por la imprecisión o vaguedad que presenta la naturaleza de las variables usadas. Por ejemplo algunas variables que se manejan en la imprecisión de su naturaleza misma son: temperatura, humedad, nivel de pobreza, entre otras.

Es de entender que la teoría de probabilidad se ha estado usando para el manejo de datos en la incertidumbre, y de esta manera se ha de plantear un enfoque propiamente estadístico, sin embargo hay que considerar una definición algo más conservadora sobre la teoría de probabilidad, la misma se trata de: “*la incertidumbre en la ocurrencia de sucesos bien definidos*”, de este modo tenemos una limitante, ya que realmente no en todos los ámbitos de estudio se puede definir un suceso de manera tan clara.

Por otro lado cuando se abarca por primera vez temas sobre: Lógica Difusa, conjuntos Difusos o Números Borrosos se producen opiniones un tanto apresuradas sobre el sustento de los mismos, algunas críticas o comentarios como el siguiente: “*La teoría borrosa es errónea, errónea y perniciosa. El peligro de la teoría borrosa es que fomenta el pen-*

¹ Difusos, borrosos o en ingles fuzzy.

samiento impreciso que nos traerá problemas²". Sin embargo el sustento de esta teoría está muy fuertemente respaldado, tanto en su soporte matemático como en su campo aplicativo a distintas áreas.

Una definición más reservada (Morillas, 2002) sobre conjuntos difusos es la siguiente: *"es el grado de ocurrencia de sucesos mal definidos"* con este enfoque mencionado por Morillas y otros autores más, nos da a entender que la incorporación de modelos que manejen datos en incertidumbre es una alternativa más, paralelamente a la definición de probabilidad anteriormente aclarada.

Cabe recalcar que tanto la Teoría de Probabilidad como la Lógica Difusa, se desarrollan de manera distinta, y no es la competencia en su aplicación la que se ha de exponer en el presente trabajo, si no por el contrario su mutuo actuar en su aplicación.

1.2. Antecedentes

Desde ya casi más de cinco siglos la teoría de la probabilidad ha dominado el campo de las matemáticas sobre la incertidumbre. Los conceptos de probabilidad datan desde el año 1500, en esa época Cardano³ reconoció las reglas de la probabilidad en los juegos de azar, y estos conceptos no fueron cuestionados sino hasta 1685 por Bishop de Wells, quien publicó un escrito, el cual indica que los estados de verdad o falsedad en la probabilidad no necesariamente deben de ser definidos en una manera tan exacta y excluyente.

1.2.1. Origen de los Conjuntos Difusos

En la antigua Grecia, las leyes del conocimiento propuestas por Aristóteles fueron las siguientes: *principio de identidad, ley de la contradicción y ley del tercero excluido*, esta última afirma que toda proposición debe ser clasificada como verdadera ó falsa, sin existir un término intermedio entre ellas. Un concepto complementario fue planteado por Platón, el cual estableció la existencia de una tercera clasificación entre el falso y verdadero, la misma se denominó como: *los grados de pertenencia*.

Posteriormente en el siglo XVIII George Berkeley y David Hume, describieron conceptos sobre lo vago, similar y la ambigüedad. Fue Jan Lukasiewicz en 1930 quien propuso por primera vez una alternativa a la lógica Aristotélica, su propuesta fue denominada como: *lógica de vaguedades*, esta fue descrita como lógica trivaluada, asumiendo al tercer valor como un *"posible"*.

²Profesor William Kahan 1975 en la presentación de discurso del Doctorado Honoris Causa por la facultad de Ciencias de la Universidad de Oviedo otorgado a Lotfy A. Zadeh, España (01-12-2005)

³Girolamo Cardano (1501- 1576) autor del primer tratado serio sobre probabilidad titulado: *"Liber de Ludo Aleae"* en 1560, pero publicado póstumamente en 1663.

El filósofo cuántico Max Black⁴ es denominado como el precursor de la lógica difusa, ya que en 1937 define el primer conjunto difuso mediante el uso de una función, la cual permitía pasar de un estado a su opuesto o mejor dicho de una falsedad a una veracidad completa. Sin embargo no fue sino hasta la década de los 60's cuando el profesor L. Zadeh, basado en las ideas de Black creó el concepto de lógica difusa, combinando los conceptos de los conjuntos de Lukasiewicz mediante la definición de grados de pertenencia, en su trabajo titulado conjuntos difusos (Zadeh, 1965).

Posteriormente de la publicación de Zadeh, le siguieron variedad de publicaciones con una mayor profundidad teórica y aplicativa, entre los cuales se tienen: (Klir & Yuan, 1995); (Dubois & Prade, 1979); (Ross, 2004), (Tanaka, Okuda, & Asai, 1976), entre otros.

Dentro de estos trabajos y otros más, se han desarrollado distintas técnicas en la aplicación de los conjuntos difusos y la lógica difusa, es el caso de los pronósticos y las series de tiempo, un área en la cual han brindado resultados muy satisfactorios.

1.2.2. Conjuntos Difusos

Antes plantear de manera más formal un conjunto difuso, es necesario mencionar la definición para los conjuntos convencionales o clásicos, la mayoría de estos últimos se puede definir generalmente de las siguientes tres maneras:

Por *enumeración* o *extensión* de los elementos que pertenecen al mismo.

$$A = \{a_1, a_2, a_3, \dots\} \quad (1.1)$$

Por *comprensión* de propiedades, la cual cumple cada elemento del mismo.

$$A = \{x \in \mathbb{R} / x \leq a_0\} \quad (1.2)$$

Por el uso de la *función de pertenencia* o *correspondencia*, denotada por $\mu_A(x)$ y definida como $\mu_A(x) : U \rightarrow \{0,1\}$, la misma descrita más concretamente como:

$$\mu_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases} \quad (1.3)$$

Cabe recalcar que el conjunto universal está denotado por U . Ahora bien la función μ_A definida en (1.3) es esencial en conjuntos difusos, ya que de ahora en adelante no únicamente poseerá como únicas imágenes a los puntos $\{0,1\}$, sino de manera más general, abarca todo el intervalo $[0; 1]$, dando así lugar a la existencia de los grados de incertidumbre.

⁴ Max Black (1909-1988) filósofo cuántico, autor del tratado titulado: “*Vagueness: An exercise in logical analysis in the Philosophical Society*”.

Definición 1.1 (L. Zadeh) Sea un espacio de puntos \mathcal{X} . Un conjunto borroso A en \mathcal{X} , se caracteriza por una función de pertenencia $\mu_A(x)$ que asocia a cada punto de \mathcal{X} un número real en el intervalo $[0,1]$, donde $\mu_A(x)$ representa el grado de pertenencia de x al conjunto difuso A :

$$A = \{(x, \mu_A(x)) : x \in \mathcal{X}, \mu_A(x) \in [0,1]\} \quad (1.4)$$

Cuanto más se aproxima $\mu_A(x)$ a la unidad, mayor es el grado de pertenencia de x al conjunto A . Con esta definición se deduce que los conjuntos clásicos por así decirlo, son un caso particular de los conjuntos difusos. Zadeh además cita lo siguiente:

“Aunque la función de pertenencia de un conjunto borroso tiene cierto parecido con una función de probabilidad, cuando X es un conjunto numerable o continuo, hay diferencias sustanciales entre ambos conceptos, de hecho la noción de conjunto borroso es de naturaleza no estadística”

Es así que en el avance del presente trabajo se expondrán conceptos y propiedades que aclararan dichas diferencias, pero a su vez se expondrá todo el aporte de la probabilidad y estadística en el área de los conjuntos difusos.

Por otro lado las representaciones de conjuntos difusos pueden variar sustancialmente, según su campo referencial⁵, dicho de este modo para un campo referencial discreto se tiene la siguiente representación de un conjunto difuso A :

$$A = a_1/u_1 + a_2/u_2 + \dots a_i/u_i + \dots a_k/u_k \quad (1.5)$$

En donde los términos $a_i \in [0; 1]$ y por lo tanto representan el grado de pertenencia de cada elemento $u_i \in \mathcal{X}$, o dicho de otra manera se tiene que $\mu_A(u_i) = a_i$. Esta representación es de mucha utilidad para casos en donde el conjunto referencial es discreto, y de la misma manera para establecer relaciones entre conjuntos difusos.

Ejemplo 1: Definamos el conjunto difuso denotado por A , donde representa a los números positivos y grandes, sobre el conjunto referencial \mathbb{R} y asignando la siguiente función de pertenencia:

$$\mu_A(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-x/100} & \text{si } x > 0 \end{cases}$$

Este conjunto claramente es preciso en el sentido de ser positivo, sin embargo en la definición de número grande no lo es, ya que el término grande está sujeto a mucha ambi-

⁵ Conjunto referencial, campo referencial o espacio de discusión, algunos autores asemejan en cierta manera al dominio de una variable aleatoria.

güedad. Esta función denotada por μ_A , si bien es la de distribución para un modelo exponencial, pero el significado de esta última es muy distinta en comparación a la de una función de pertenencia.

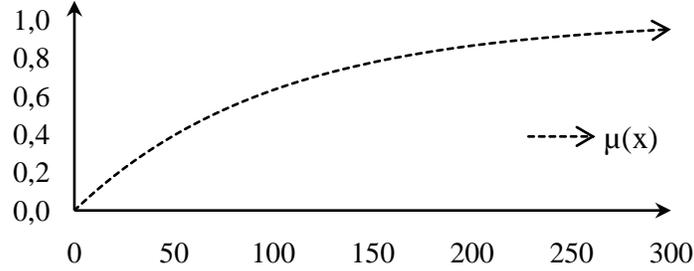


Ilustración 1. Representación de la función de pertenencia: $1 - e^{-x/100}$

De la misma manera que en los conjuntos convencionales, en los conjuntos difusos se definen los operadores unarios y binarios, pero en estos últimos el manejo de dichos operadores está en función a $\mu(x)$, es decir son las funciones de pertenencia de los conjuntos difusos las que son tratadas de manera más explícita.

El *complemento* de un conjunto difuso A , se representa por \bar{A} y su función de pertenencia se define como:

$$\bar{A} \rightarrow \mu_{\bar{A}}(x) = 1 - \mu_A(x) \quad (1.6)$$

Sean dos conjuntos A y B en X , con funciones $\mu_A(x)$ y $\mu_B(x)$ respectivamente, para la *intersección* denotada por $A \cap B$, se define la función de pertenencia como:

$$A \cap B \rightarrow \mu_{A \cap B}(x) = \mu_A(x) \wedge \mu_B(x) = \min(\mu_A(x), \mu_B(x)) \quad (1.7)$$

La *unión* de los conjuntos A y B en X , denotada por $A \cup B$, define su función de pertenencia de la siguiente manera:

$$A \cup B \rightarrow \mu_{A \cup B}(x) = \mu_A(x) \vee \mu_B(x) = \max(\mu_A(x), \mu_B(x)) \quad (1.8)$$

Dado el carácter introductorio del actual capítulo, no se abarcaran por ahora otros operadores o propiedades de los conjuntos difusos. Por otro lado cabe recalcar el siguiente hecho, de que el conjunto $A \cap \bar{A}$ no siempre es el *conjunto vacío* \emptyset ya que este último definido como:

$$\emptyset = \{ \forall x \in X : \mu_{\emptyset}(x) = 0 \} \quad (1.9)$$

Claramente se puede verificar que no siempre se cumple este hecho ($A \cap \bar{A} \neq \emptyset$), debido a la definición de intersección (1.6) y la de complemento (1.5), de manera similar se puede verificar la siguiente relación $A \cup \bar{A} \neq U$ (donde U es el conjunto universal), en base a la definición de unión (1.7).

Ejemplo 2: Definamos los conjuntos difusos A : “Aproximadamente 5” y B : “Aproximadamente 6”. A se representan de manera gráfica a los conjuntos Conjunto $A, B, A \cup B$ y $A \cap B$ respectivamente.

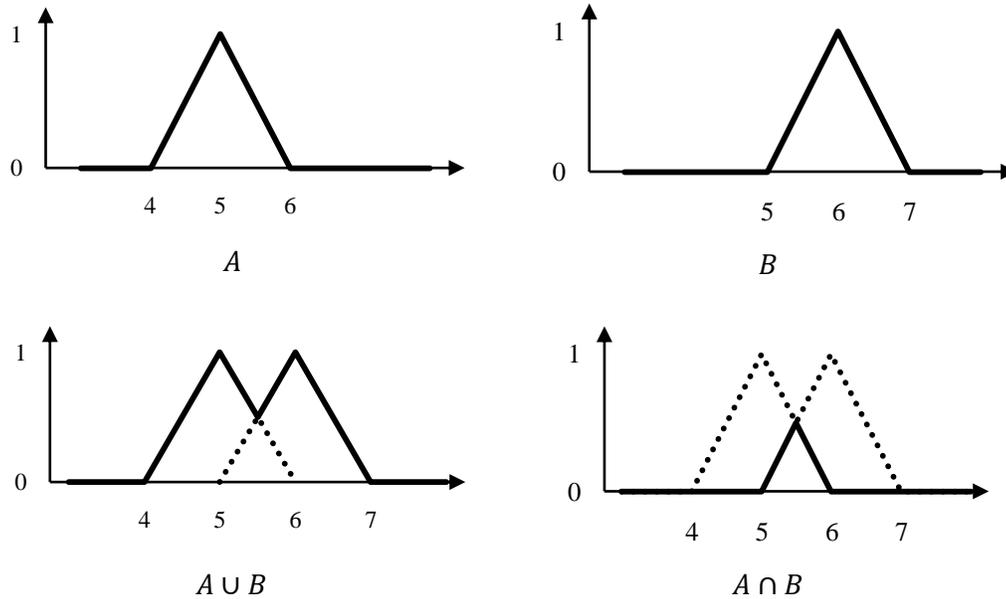


Ilustración 2. Diferentes operaciones: $A, B, A \cup B$ y $A \cap B$

Una descripción y clasificación más general de los operadores entre conjuntos difusos se encuentra en el trabajo de Ross T. en el mismo se detallan los denominados compensadores, t-normas y t-c-normas. Por otro lado una definición importante para la descomposición de conjuntos difusos y para su manejo es la siguiente.

Definición 1.2 Sea un conjunto difuso A en \mathcal{X} , y sea un número $\alpha \in [0,1]$, se define el α – cut y el strong α – cut del conjunto difuso, A_α y $A_{\alpha+}$ respectivamente como:

$$A_\alpha = \{x \in \mathcal{X}, \mu_A(x) \geq \alpha\} \quad (1.9)$$

$$A_{\alpha+} = \{x \in \mathcal{X}, \mu_A(x) > \alpha\} \quad (1.10)$$

Esta última definición es muy esencial, ya que brinda una flexibilidad en el manejo de operadores entre conjuntos difusos, lo cuales abarcara conforme al avance del trabajo en sí mismo. Una representación gráfica para los α – cut se muestra en la Ilustración 3.

Los α – cut o alfa-cortes, describen los distintos niveles de seguridad o posibilidad en un conjunto difuso, y de la misma manera determinan ciertas propiedades que deben cumplir los conjuntos difusos a ser manejados.

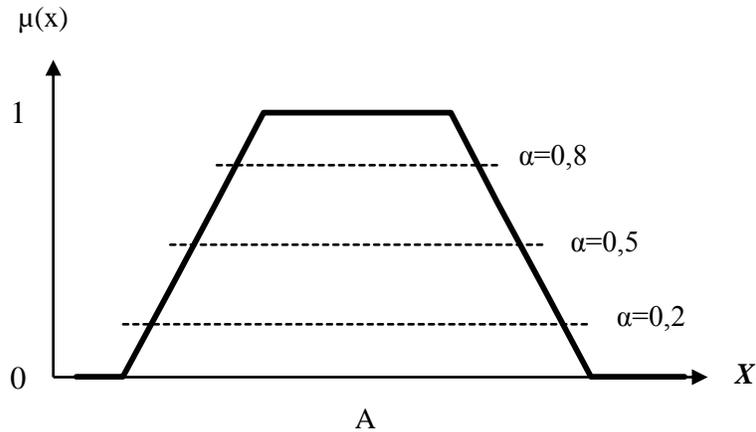
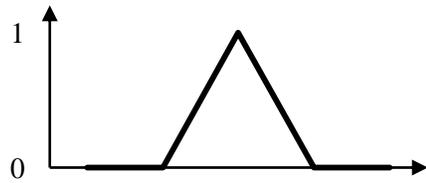
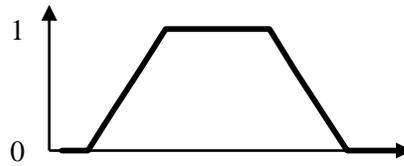


Ilustración 3. Representación grafica de alfa-cortes

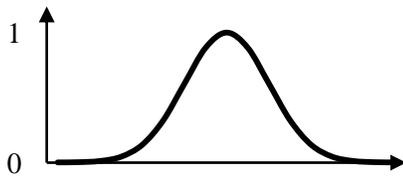
Cuando se habla de un conjunto difuso, lo primero que se debe de pensar es en definir es la función de pertenencia del mismo, es así que la mayoría de los autores catalogan a los conjuntos difusos según el tipo de función de pertenencia que poseen ó que se le puede asignar. Entre dichas funciones las comúnmente usadas son las siguientes: Funciones triangulares, trapezoidales, gaussianas y las lingüísticas.



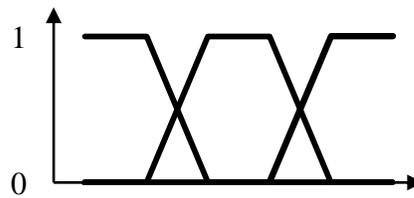
Triangular



Trapezoidal



Gausiana



Lingüística

Ilustración 4. Representación de diferentes funciones de pertenencia

Dado que en la teoría difusa también se pueden considerar la modelación de sentencias del lenguaje natural o aspectos imprecisos del mundo real a un lenguaje matemático formal, es necesario el manejo de variables lingüísticas.

1.2.3. Variables Lingüísticas

Es una variable donde los valores son palabras u oraciones en lenguaje natural, como por ejemplo, la variable lingüística *calidad*, y sus valores posibles pueden ser *baja*, *media* y *alta*.

Definición 1.3 (I. Álvarez) Una variable lingüística se caracteriza mediante la estructura $\langle X, L_\theta, \theta, M_\theta \rangle$; donde X es el nombre de la variable, L_θ es el conjunto de valores lingüísticos, θ es el dominio físico de la variable, M_θ es la función semántica que asigna a los valores lingüísticos de la variable.

Las variables lingüísticas generalmente se usan para representación de conjuntos difusos, los cuales a su vez definen ciertas expresiones de cambios, como por ejemplo:

- A_1 : Incremento Muy Elevado.
- A_2 : Incremento Elevado.
- A_3 : Incremento.
- A_4 : Incremento Leve.
- A_5 : Incremento Bajo.

En estas 5 expresiones lingüísticas, se complementan con las de decremento, y así contemplan todo el conjunto de valores lingüísticos, sobre la variable “*situación de cambio para una serie dada*”, y son estas expresiones lingüísticas que se usan para las series de tiempo difusas, y para un mejor manejo de interpretación de los datos.

1.2.4. Serie de Tiempo Difusa

En un principio Zadeh no propone o describe alguna relación entre los conjuntos difusos y su aplicación en el campo de las series de tiempo. Fueron Song & Chissom’s (Song & Chissom, 1993), quienes plantearon de manera más formal el concepto de serie de tiempo difusa, considerándolos de esta manera como los creadores de esta teoría⁶.

Definición 1.4 (Song & Chissom’s) Sea $Y(t)$ con $(t = 0, 1, 2, \dots)$, un subconjunto de \mathbb{R} , siendo el universo de discurso \mathcal{X} , para los conjuntos difusos $f_i(t): i = 1, 2, 3, \dots$. Si $F(t)$ es la colección de $f_i(t)$, entonces $F(t)$ se denomina Serie de Tiempo Difusa sobre $Y(t)$.

Cuando Song & Chissom’s plantearon la definición de una serie de tiempo difusa, igualmente se desarrollaron las relaciones referentes a dicha serie, estas mismas se basan en la relación de un antes y un después. Tanto las relaciones y los operadores binarios entre conjuntos difusos se describen más ampliamente en otras publicaciones (Klir & Yuan, 1995), estos operadores dieron el soporte para que Song & Chissom’s desarrollen su uso en el campo de series de tiempo difusas, de una manera más específica.

⁶En este sentido fueron Song y Chissom’s los creadores de la teoría sobre series de tiempo difusas.

Dichas relaciones dieron inicio a una serie de nuevos conceptos sobre las mismas, una de las principales se plasma en la siguiente definición:

Definición 1.5 *Suponga que $F(t)$ es causado solamente por $F(t - 1)$, mediante la relación $F(t) = F(t - 1) \times R(t; t - 1)$. Si para todo t el término R es independiente de t entonces $F(t)$ se denomina serie de tiempo difusa-invariante, de lo contrario se considerara serie de tiempo difusa-variante.*

Es de entender que tras la teoría planteada por Song & Chissom's, se propusieron paralelamente algoritmos para el análisis de las series de tiempo difusas. Sin embargo, se desarrollo un nuevo aporte (Chen, 1996) tanto la parte teórica como la parte metodológica plasmándola en un nuevo modelo o como se denominan *reglas de Chen* en base STD, poniendo mayor facilidad en su manejo.

Dentro de las distintas metodologías usadas en las series difusas, ya sea para el caso variante como invariante, la mayoría se centra en confeccionar la función de relación denotada por R . El aporte esencial por Chen fue el de simplificar la metodología sobre el manejo de las series, esta simplificación se debe a la incorporación de *promedios o agregadores* dentro de la confección de los pronósticos.

Brinda mucha desenvoltura el uso de las reglas de Chen, pero Sullivan & Woodall (Sullivan & Woodall, 1994) publicarían su artículo en el cual se maneja la *matriz de transición* para una cadena de Markov, y con esto realizar los pronósticos usando solo el concepto de Song & Chissom's.

El proceso de estimación de matriz de transición de una cadena de Markov se desarrollo específicamente por Anderson (Anderson & Goodman, 1955), todo esto bajo los conceptos clásicos de la función de *máxima verosimilitud* y el uso de muestras multinomiales, sin embargo la agregación de técnicas bayesianas en la estimación se desarrolla posteriormente.

La inferencia bayesiana por otro lado considera varios criterios para la obtención de estimadores adecuados, Strelloff & Crutchfield (1995) se enfocan en la incorporación de la *distribución Dirichlet* en la fase de estimación para datos multinomiales, mejorando considerablemente lo sostenido por Anderson & Goodman.

Dado lo expuesto en los párrafos anteriores, el enlace que une el uso de promedios de Chen y la distribución Dirichlet en la estimación de una cadena de Markov, es el concepto de: "*variable aleatoria difusa*" incorporado por Puri & Ralescu (1986), en donde se asigna probabilidades a variables aleatorias difusas, pero su manejo es estrictamente en el campo de números difusos.

1.3 Planteamiento del problema

Tal como se mencionó al inicio, la exactitud en la medición de las variables a ser analizadas es una delimitación inicial, a pesar de esto, la mayoría de las estimaciones o pronósticos por métodos probabilísticos proporcionan intervalos de confianza para sus estimaciones, pero no es la naturaleza del presente trabajo discutir esta metodología.

La problemática en sí, es la incorporación de nuevas técnicas de pronósticos o mejor dicho la introducción de conceptos nuevos los cuales son: *variables aleatorias difusas* y *la distribución Dirichlet* en las series de tiempo difusas.

Más concretamente, el centro de investigación del presente trabajo gira en la siguiente pregunta: *¿la introducción del modelo de Markov difuso brinda una mejor ajuste en comparación a los métodos usuales de series de tiempo?*.

1.4. Hipótesis

Es posible generar una metodología alternativa de pronósticos de series de tiempo univariado en base al Modelo de Markov y la teoría difusa.

1.5. Objetivos

1.5.1. Objetivo General

Detallar el proceso que envuelve las técnicas de pronósticos en series de tiempo difusas, específicamente el modelo de Markov difuso.

1.5.2. Objetivos Específicos

- Describir los procedimientos sobre el manejo de conjuntos difusos, números difusos y *variables aleatorias difusas*.
- Exponer los conceptos necesarios para el uso de la *distribución Dirichlet*, en la estimación bayesiana de parámetros multinomiales dentro de la matriz de transición de Markov.
- Comparar la metodología del *Modelo de Markov Difuso* propuesto, con la métodos usuales de manejo en series de tiempo, más concretamente con la metodología ARIMA de *Box & Jenkins*⁷.

⁷ Box, & Jenkins “*Time Series Analysis: Forecasting and Control*” 1ra edición-1970.

1.6. Justificación

La mayoría de las técnicas de pronósticos en series de tiempo, necesitan ciertos supuestos o asunciones. Muchos supuestos y afirmaciones que se manejan son los siguientes: normalidad, estacionareidad, heterosedasticidad, entre otros.

A diferencia de esto las series de tiempo difusas no se ven necesarios estos supuestos. Sin embargo la falta de una formalidad estadística en el manejo de estos métodos es una limitante, sobre todo en la línea de trabajo planteada por Chen, donde se hace su uso poco justificado de las medias generalizadas, dando hincapié al desarrollo de este trabajo.

El acelerado avance de la teoría fuzzy o difusa en varias ramas, hace necesario el conocer sus fundamentos y su mutuo actuar con la estadística, esto mediante el manejo de números difusos y variables aleatorias difusas.

Sin embargo cabe destacar que es el investigador o usuario quien decide que técnica o metodología debe de aplicar.

1.7. Alcances y limitaciones

Si bien toda metodología posee tanto alcances como limitaciones, la teoría de series de tiempo difusas de la misma manera presenta los siguientes aspectos:

- Un manejo univariado de datos y sus pronósticos respectivos.
- Manejo de incertidumbre en la medición de datos.
- Mayor rapidez en el cálculo.
- Manejo de datos incompletos⁸ o exclusión de datos atípicos.
- Ausencia de supuestos necesarios⁹.

⁸ Se puede trabajar con series de datos incompletas, ya que no necesita continuidad en la medición de datos cronológicos.

⁹ En este sentido se referirá a los supuestos como son: normalidad, heterosedasticidad, entre otros.

Capítulo II

Marco Teórico

2.1. Conjuntos difusos

En el presente capítulo abarcaremos el respaldo de la lógica difusa y por ende también el de las series de tiempo difusas, este respaldo son los ya mencionados conjuntos difusos implementados por Lofty Zadeh (Zadeh, 1965). Además se han de trazar los lineamientos necesarios que relacionan su manejo y requisitos o propiedades que los mismos poseen.



Ilustración 5. Lofty Zadeh el denominado padre de la teoría fuzzy

Cabe aclarar que solo se abarcara conceptos mínimos para el tema principal de series de tiempo difusas, y que escapan conceptos más especializados sobre lógica difusa, el cual se considera el área de mayor manejo en la teoría de conjuntos difusos.

Para el primer capítulo se vio definiciones y algunos ejemplos sobre conjunto difusos, además de los operadores para intersección, unión y complemento. Ahora se detallaran algunas definiciones y propiedades habituales sobre los conjuntos difusos de una manera más general.

Definición 2.1 (Cardinalidad) La Cardinalidad de un conjunto difuso A en \mathcal{X} , se define como:

$$|A| = \begin{cases} \sum_i \mu_A(x_i) & \text{cuando } \mathcal{X} \text{ es Discreto} \\ \int_{-\infty}^{+\infty} \mu_A(x) dx & \text{cuando } \mathcal{X} \text{ es Continuo} \end{cases} \quad (2.1)$$

Definición 2.2 (Altura) La Altura de un conjunto difuso A en \mathcal{X} denotada por $High(A)$, define como: $High(A) = Max(\mu_A(x_i))$, es decir el valor máximo de la función de pertenencia μ_A .

La utilidad de estas definiciones y otras más que se presenten, junto a los resultados obtenidos a través de ellas, se harán más evidentes conforme se dé avance al tema de discusión.

Definición 2.3 (Soporte) El soporte de un conjunto difuso A en \mathcal{X} , consiste en todos los elementos con valor de pertenencia mayor a 0, es decir:

$$Supp(A) = \{x \in \mathcal{X} / \mu_A(x) > 0\} \quad (2.2)$$

Definición 2.4 (Normalidad) Un conjunto difuso A en \mathcal{X} , se considera Normal, si existe al menos un elemento de certeza segura, es decir:

$$A \text{ es Normal} \Leftrightarrow \exists x \in \mathcal{X} / \mu_A(x) = 1 \quad (2.3)$$

En caso de que esta última definición no se cumpliera, es decir un conjunto no es Normal, entonces el conjunto se denominara como conjunto *Subnormal*.

Se pone a conocimiento la existencia de más definiciones que acompañan a un conjunto difuso, sin embargo estas no se consideran muy relevantes, dado el carácter del enfoque, para tener un detalle más profundo sobre algunas de estas, se recomiendan trabajos más específicos.

Se expusieron la función de pertenencia, y antes de implicar los distintos tipos de funciones de pertenencia que poseen los conjuntos difusos, es necesaria la siguiente definición:

Definición 2.5 (Convexidad) Un conjunto difuso A en \mathcal{X} , se denominara convexo si:

$$\mu_A(\lambda x_1 + (1 - \lambda) x_2) \geq Min(\mu_A(x_1); \mu_A(x_2)) \quad \forall x_1, x_2 \in \mathcal{X} \text{ y } \forall \alpha \in [0; 1] \quad (2.4)$$

De la misma manera que en caso de la normalidad, una función que no cumpla esta propiedad ha de definir a un conjunto como *no convexo*. Consecuentemente los α - cut de un conjunto difuso, deben de ser convexos si la función μ_A lo es.

2.1.1. Funciones de pertenencia

Cuando se maneja un conjunto difuso lo esencial es conocer la función de pertenencia, esta misma definirá al conjunto en sí. Existen infinitas funciones de pertenencia, que se usan en la teoría de conjuntos difusos, como se vio previamente, se detallaron alguna de ellas, sin embargo entre estas y otras más se destacan las que a continuación se mencionaran:

Función Triangular

$$\mu(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ \frac{c-x}{c-b} & b \leq x \leq c \\ 0 & c \leq x \end{cases} \quad (2.5)$$

Función Trapezoidal¹⁰

$$\mu(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & a \leq x \leq b \\ \frac{c-x}{c-b} & b \leq x \leq c \\ 0 & c \leq x \end{cases} \quad (2.6)$$

Función Singleton

$$\mu(x) = \begin{cases} 1 & x = a \\ 0 & x \neq a \end{cases} \quad (2.7)$$

Función Gaussiana

$$\mu(x) = \begin{cases} e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \end{cases} \quad (2.8)$$

Función de Bell

$$\mu(x) = \begin{cases} \frac{1}{1 + \left|\frac{x-c}{c}\right|^{2b}} \end{cases} \quad (2.9)$$

Función Sigmoide (o sigmoidal)

$$\mu(x) = \begin{cases} \frac{1}{1 + e^{-a(x-c)}} \end{cases} \quad (2.10)$$

Función Left-Right (izquierda-derecha)

$$\mu(x) = \begin{cases} F_L\left(\frac{c-x}{\alpha}\right); & x < c \\ F_R\left(\frac{x-c}{\beta}\right); & x \geq c \end{cases} \quad (2.11)$$

Es esta última función denominada LR, la cual varios autores consideran la generalización de las anteriores debido a su expresión más universal, por otro lado son las fun-

¹⁰ Esta función es esencial en la investigación, ya que es la más usada de las series de tiempo difusas

ciones (2.5) y (2.6) las que preferentemente se usan en el tema de FTS, sin restar crédito del manejo a las anteriores dado su fácil planteo en la aplicación.

2.1.2. Operadores entre Conjuntos Difusos

Para el primer capítulo se mencionó algunos operadores unarios y binarios sobre conjuntos difusos, algunos de ellos fueron el complemento y la unión respectivamente hablando, cada uno de estos opera uno o más conjuntos difusos en términos de sus funciones de pertenencia.

Ahora bien se ha de enfatizar y generalizar más, sobre algunos de estos operadores, pero antes es necesario especificar dos aspectos importantes en esta parte.

Definición 2.6 (T-norma) *Es una función bivalente denotada por "t", la cual tiene como dominio ha $t: [0; 1] \times [0; 1] \rightarrow [0; 1]$ y este operador mismo, tiene que satisfacer las siguientes propiedades:*

- $t(0; \alpha) = t(\alpha; 0) = 0 \quad \forall \alpha \in [0; 1]$
- $t(1; \alpha) = t(\alpha; 1) = \alpha \quad \forall \alpha \in [0; 1]$
- Si $\alpha \leq \gamma$ y $\beta \leq \delta \Rightarrow t(\alpha; \beta) \leq t(\gamma; \delta)$
- $t(\beta; \alpha) = t(\alpha; \beta) \quad \forall \alpha, \beta \in [0; 1]$
- $t(\alpha; t(\beta; \gamma)) = t(t(\alpha; \beta); \gamma) \quad \forall \alpha, \beta, \gamma \in [0; 1]$

También denominadas normas triangulares, la t-norma se usa como un operador de intersección. En el capítulo anterior se vio al operador de intersección bajo la t-norma del mínimo entre α y β , el mismo cumple todas las propiedades antes mencionadas.

A continuación se detallan algunas de los tipos de estos operadores de intersección (t-norma) que se manejan generalmente:

Dombi [1982]

$$\left\{ 1 + \left[\left(\frac{1}{\alpha} - 1 \right)^\lambda + \left(\frac{1}{\beta} - 1 \right)^\lambda \right]^{\frac{1}{\lambda}} \right\}^{-1}$$

Frank [1979]

$$\log_s \left[1 + \frac{(s^\alpha - 1)(s^\beta - 1)}{s - 1} \right]$$

Hamacher [1978]

$$\frac{\alpha\beta}{r + (1 - r)(\alpha + \beta - \alpha\beta)}$$

Schweizer & Sklar(I) [1963]

$$[\max(0; \alpha^p + \beta^p - 1)]^{\frac{1}{p}}$$

Schweizer & Sklar (II) [1963]

$$1 - [(1 - \alpha)^p + (1 - \beta)^p - (1 - \alpha)^p(1 - \beta)^p]^{\frac{1}{p}}$$

Schweizer & Sklar (III) [1963]

$$\exp \left[-(|\ln(\alpha)|^p + |\ln(\beta)|^p)^{\frac{1}{p}} \right]$$

Schweizer & Sklar (IV) [1963]

$$\frac{\alpha\beta}{[\alpha^p + \beta^p - \alpha^p\beta^p]^{\frac{1}{p}}}$$

Yager [1980]

$$1 - \min \left\{ 1; [(1 - \alpha)^w + (1 - \beta)^w]^{\frac{1}{w}} \right\}$$

Dubois & Prade [1980]

$$\frac{\alpha\beta}{\max(\alpha; \beta; w)}$$

Weber [1983]

$$\max \left(0; \frac{\alpha + \beta + \alpha\beta w - 1}{1 + w} \right)$$

Tal como se puede apreciar, en una mayoría de estas normas ostentan parámetros los cuales deben de poseer un valor numérico para poder operar con los mismos, algunos poseen semejanza entre ellos al suponer valores límites para ellos¹¹.

Definición 2.7 (S-norma) Es una función bivalente denotada por "s", la cual tiene como dominio ha $s: [0; 1] \times [0; 1] \rightarrow [0; 1]$ y este operador mismo, tiene que satisfacer las siguientes propiedades:

- $s(0; \alpha) = s(\alpha; 0) = \alpha \quad \forall \alpha \in [0; 1]$
- $s(1; \alpha) = s(\alpha; 1) = \alpha \quad \forall \alpha \in [0; 1]$
- Si $\alpha \leq \gamma$ y $\beta \leq \delta \quad \Rightarrow \quad s(\alpha; \beta) \leq s(\gamma; \delta)$
- $s(\beta; \alpha) = s(\alpha; \beta) \quad \forall \alpha, \beta \in [0; 1]$
- $s(\alpha; s(\beta; \gamma)) = s(s(\alpha; \beta); \gamma) \quad \forall \alpha, \beta, \gamma \in [0; 1]$

¹¹ En anexos se detallan más equivalencias entre las distintas t-normas presentadas.

Algunos autores también las denominan t-cnormas, y son estos operadores binarios los cuales cumplen la función de unión entre conjuntos, la función $\max(\alpha; \beta)$ es la que se planteó al inicio, siendo esta la mayor facilidad en su manejo.

De la misma manera que en las t-normas, en la s-normas conforme se desarrolló la teoría difusa se plantearon distintas alternativas; entre las más relevantes se tienen a:

Dombi [1982]

$$\left\{ 1 + \left[\left(\frac{1}{\alpha} - 1 \right)^\lambda + \left(\frac{1}{\beta} - 1 \right)^\lambda \right]^{-\frac{1}{\lambda}} \right\}^{-1}$$

Frank [1979]

$$1 - \log_s \left[1 + \frac{(s^{1-\alpha} - 1)(s^{1-\beta} - 1)}{s - 1} \right]$$

Hamacher [1978]

$$\frac{\alpha + \beta + (r - 2)\alpha\beta}{r + (r - 1)\alpha\beta}$$

Schweizer & Sklar (I) [1963]

$$1 - \{ \max(0; (1 - \alpha)^p + (1 - \beta)^p - 1) \}^{\frac{1}{p}}$$

Schweizer & Sklar (II) [1963]

$$[\alpha^p + \beta^p - \alpha^p \beta^p]^{\frac{1}{p}}$$

Schweizer & Sklar (III) [1963]

$$1 - \exp \left[- \left(|\ln(1 - \alpha)|^p + |\ln(1 - \beta)|^p \right)^{\frac{1}{p}} \right]$$

Schweizer & Sklar (IV) [1963]

$$\frac{(1 - \alpha)(1 - \beta)}{[(1 - \alpha)^p + (1 - \beta)^p - (1 - \alpha)^p (1 - \beta)^p]^{\frac{1}{p}}}$$

Yager [1980]

$$1 - \min \left\{ 1; [\alpha^w + \beta^w]^{\frac{1}{w}} \right\}$$

Dubois & Prade [1980]

$$\frac{(1 - \alpha)(1 - \beta)}{\max((1 - \alpha); (1 - \beta); w)}$$

Weber [1983]

$$\min\left(1; \alpha + \beta - \frac{w}{1-w}\alpha\beta\right)$$

Yu [1985]

$$\min(1; \alpha + \beta + w\alpha\beta)$$

Dado que lo se está tratando son conjuntos, y en la teoría acostumbrada sobre conjuntos se tienen distintas propiedades, en este sentido las s-normas y t-normas también verifican las leyes de Morgan respecto a la intersección y unión:

$$t(\alpha; \beta) = 1 - s(1 - \alpha; 1 - \beta) \quad (2.13)$$

Es decir que para cualquier s-norma denotada por "s" que cumple las propiedades de s-norma, siempre será posible hallar una t-norma "t", la cual también cumplirá las propiedades de una t-norma, esta propiedad también se conoce como el principio de dualidad¹².

Una gran cantidad de publicaciones distinguen diferentes propiedades y metodologías para la elección de una t-norma adecuada, y del cálculo de los parámetros necesarios para la misma.

Relaciones Difusas

Al igual que conceptos como t-norma y s-norma, los cuales juegan un papel importante como unión e intersección. Para conjuntos difusos una noción importante es la de una *Relación* denotada ahora y en adelante por "R" ya que estas son un aspecto trascendental para las series de tiempo Fuzzy, como se vio en las definiciones del apartado anterior, y es aquí en donde se pondrá mayor énfasis.

Antes de extender el concepto de relación a un espacio difuso, tomaremos un ámbito matemático lo más sencillo posible. Se dice que una relación binaria "R" es definida como un subconjunto del producto cartesiano de Ω (que representa al universo) es decir: $R \subseteq \Omega \times \Omega$, si se da a un par ordenado de elementos de R expresados por (a, b) , se puede determinar si está o no relacionado "a" con "b", el concepto de relación implica la idea de enumeración, de algunos de estos elementos, de conjuntos que forman tuplas y en el caso de una relación binaria, duplas.

Una relación difusa de conjuntos difusos del producto cartesiano formado por los espacios referenciales: $\mathcal{X} \times \mathcal{Y}$, es un mapeo de la siguiente manera $\mathcal{X} \rightarrow \mathcal{Y}$. Este mapeo de elementos del de \mathcal{X} hacia \mathcal{Y} a el intervalo $[0; 1]$ donde la fuerza de la relación se expresa mediante la función de asociación de pares ordenados de los dos o expresa como $\mu_R(x, y)$. Esto se puede expresar como:

¹² Del latín dualitas, el término dualidad señala la existencia de dos fenómenos o caracteres diferentes en una misma persona o en un mismo estado de cosas.

$$R = \{(x, y), \mu_R(x, y)\} / (x, y) \in \mathcal{X} \times \mathcal{Y}\}$$

Usualmente las funciones que se utilizan para operar como relaciones, son los operadores Min-Max, y las composiciones entre funciones se manejan de la misma manera que las t-normas y s-normas.

Al iniciar el avance teórico por Song & Chissom's sobre las series de tiempo fuzzy, estas relaciones son un papel importante en todo el proceso de análisis.

2.1.3. Números Difusos

Hasta ahora se han detallado varias propiedades básicas sobre conjuntos difusos, pero el siguiente concepto lleva la teoría difusa de conjuntos, a una semejanza comparable al manejo de los números en el campo complejo, datos composicionales en el espacio simplex¹³, entre otros por mencionar.

Definición 2.8. (Numero difuso) *Un conjunto difuso A en X, con función de pertenencia denotada por μ_A , se denominara número difuso si es normal en solo un punto $x_{(m)}$, y es convexo sobre su función μ_A .*

En este sentido un número difuso es una prolongación de un número conocido en el sentido que no se refiere a un único valor sino a un agregado de posibles valores, que varían con un peso entre 0 y 1. Un número difuso es así un caso especial de conjunto difuso convexo.

Así como la lógica difusa es una extensión de la lógica booleana, los números difusos son una extensión de los números reales, con lo cual siempre se ha de entender que los números difusos posean el espacio referencial X a los números reales R. Los cálculos con números difusos permiten la incorporación de incertidumbre en parámetros, propiedades, geometría, condiciones iniciales, entre otras más.

Operaciones Aritméticas entre Números Difusos

Inauguralmente se vio el manejo entre dos conjuntos difusos cualesquiera A y B, las operaciones elementales expuestas fueron las siguientes:

- Complemento $\bar{A} \rightarrow \mu_{\bar{A}}(x) = 1 - \mu_A(x)$
- Intersección $A \cap B \rightarrow \mu_{A \cap B}(x) = \mu_A(x) \wedge \mu_B(x) = \min(\mu_A(x), \mu_B(x))$
- Unión $A \cup B \rightarrow \mu_{A \cup B}(x) = \mu_A(x) \vee \mu_B(x) = \max(\mu_A(x), \mu_B(x))$

¹³ El espacio simplex es un subconjunto del espacio k-dimensional de los números reales, en donde las k componentes suman una constante única, para cualquier elemento del espacio denominado simplex.

Pero para un número difuso estas operaciones se extienden a otras más por las características más flexibles que ofrecen, esto sin dejar de lado el concepto de t-norma “ \wedge ” y s-norma “ \vee ”. Dicho esto, sean \tilde{U} y \tilde{W} dos números difusos en \mathbb{R} , se definen las cuatro operaciones binarias sobre estos como:

- Suma

$$(\tilde{U} + \tilde{W})_{(z)} = \sup_{x+y=t} (\tilde{U}_{(x)} \wedge \tilde{W}_{(y)})$$

- Resta

$$(\tilde{U} - \tilde{W})_{(z)} = \sup_{x-y=t} (\tilde{U}_{(x)} \wedge \tilde{W}_{(y)})$$

- Producto

$$(\tilde{U} * \tilde{W})_{(z)} = \sup_{x*y=t} (\tilde{U}_{(x)} \wedge \tilde{W}_{(y)})$$

- División

$$\left(\frac{\tilde{U}}{\tilde{W}}\right)_{(z)} = \sup_{x=t*y} (\tilde{U}_{(x)} \wedge \tilde{W}_{(y)})$$

Universalmente existen infinidad de operaciones más que se pueden exponer de los números reales para desarrollarlos en el campo difuso, sin embargo enfatizaremos en los números difusos que posean o manejen la función de tipo trapezoidal (2.6).

Kaufmann & Aluja señalan concretamente: “*Con la utilización de números difusos triangulares (si es necesario los números L-R de Dubois & Prade que los generalizan) el experto puede establecer las correspondientes operaciones en lo incierto, en el supuesto de realizaciones que deben encadenarse en el tiempo. Se recomienda la utilización de números difusos triangulares por su simplicidad y buena percepción por parte de los que no son matemáticos (...). Este instrumento es simple y se adapta bien a los medios de información*”.

Basándonos en el anterior párrafo citado, mostraremos de manera inmediata las operaciones aritméticas para los números difusos trapezoidales, así que tomamos de nuevo dos números \tilde{U} y \tilde{W} pero en términos de cuartetos para cada uno: $\tilde{U} = (u_1, u_2, u_3, u_4)$ y $\tilde{W} = (w_1, w_2, w_3, w_4)$, además de suponer que el número resultante de la operación binaria entre estos está dado por: $\tilde{Z} = (z_1, z_2, z_3, z_4)$ como se ve en la Ilustración 6.

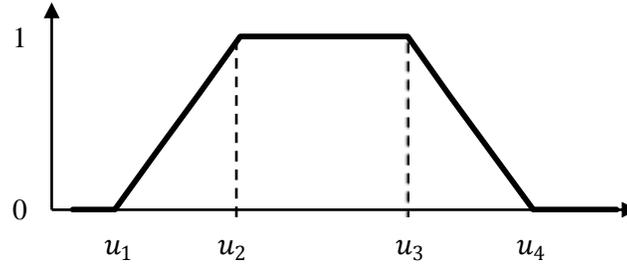


Ilustración 6. Número difuso trapezoidal en términos de cuartetas

- Suma

$$\tilde{U} + \tilde{W} = \tilde{Z} = (u_1 + w_1, u_2 + w_2, u_3 + w_3, u_4 + w_4)$$

- Resta

$$\tilde{U} - \tilde{W} = \tilde{Z} = (u_1 - w_4, u_2 - w_3, u_3 - w_2, u_4 - w_1)$$

- Producto¹⁴

$$\tilde{U} * \tilde{W} = \tilde{Z} = (u_1 * w_1, u_2 * w_2, u_3 * w_3, u_4 * w_4)$$

- División¹⁵

$$\frac{\tilde{U}}{\tilde{W}} = \tilde{Z} = \left(\frac{u_1}{w_4}, \frac{u_2}{w_3}, \frac{u_3}{w_2}, \frac{u_4}{w_1} \right)$$

Una operación unaria importante, es la de multiplicación por un escalar $k \in \mathbb{R}$, resulta en $k\tilde{U} = (ku_1, ku_2, ku_3, ku_4)$ si $k > 0$, $k\tilde{U} = (ku_4, ku_3, ku_2, ku_1)$ si $k < 0$ y cuando $k = 0$ se obtiene el elemento nulo $k\tilde{U} = (0, 0, 0, 0)$.

A pesar de que nos enfocamos en un tipo particular de números difusos, por el momento nos dispersaremos de conceptos y fundamentos más avanzados, que totalmente son muy importantes en el campo de números difusos como son: distancias, ecuaciones, derivadas, integrales, entre otros más, ya que el desarrollo de esto escapa a los alcances del presente documento.

2.1.4. Variables Aleatorias Difusas

Si bien se definió en el apartado anterior el concepto de números difusos, es algo apresurado introducir los conceptos de variables aleatorias en el campo difuso, y el darle una formalidad muy fundamentada como se hace en probabilidad clásica (Feller, 1968), estas aclaraciones se reforzaran en anexos.

¹⁴ El número difuso exacto que resulta no es un número trapezoidal, sin embargo se muestra la aproximación que Kaufman (1986) aplica bajo condiciones del $Supp(\tilde{U})$, $Supp(\tilde{W})$ y otras más.

¹⁵ Ídem a ⁽¹⁴⁾.

Desde la introducción de la teoría de conjuntos difusos, se han dedicado diversos trabajos al análisis de su conexión con la teoría de la probabilidad. Entre ellos cabe destacar los que examinan las vinculaciones de los conjuntos difusos con conceptos probabilísticos y los que establecen y manejan conceptos que combinan nociones de ambas teorías. Aclarado lo anterior, el objetivo de este apartado va a ser el desarrollo de la borrosidad que aparece en un experimento aleatorio.

En un inicio se introduce la noción de una variable aleatoria difusa (Kwakernaak, 1978), como una función $F: \Omega \rightarrow \bar{\mathcal{F}}(\mathbb{R})$ (sujeta a ciertas condiciones de medida¹⁶), donde (Ω, \mathcal{A}, P) es un espacio de probabilidad, y $\bar{\mathcal{F}}(\mathbb{R})$ denota un conjunto de varias funciones continuas por tramos $u \in [0; 1]$. A pesar de que se desarrollaron otros enfoques de variables aleatorias difusas, se expone una conceptualización matemática más operable y completa (Puri & Ralescu, 1986), con plataforma al trabajo de Kwakernaak.

Sea $(\Omega, \mathcal{A}, \mathcal{P})$ un espacio de probabilidad medible, sea $\bar{\mathcal{F}}_0(\mathbb{R}^n)$ un conjunto de elementos $u: \mathbb{R}^n \rightarrow [0; 1]$ con las siguientes propiedades:

- $\{x \in \mathbb{R}^n: u(x) \geq \alpha\}$ son conjuntos compactos¹⁷ para cada $\alpha > 0$.
- $\{x \in \mathbb{R}^n: u(x) = 1\} \neq \emptyset$.

Definición 2.9. (Puri & Ralescu) *Una variable aleatoria difusa (o variable difusa) es una función $X: \Omega \rightarrow \bar{\mathcal{F}}_0(\mathbb{R}^n)$ tal que:*

$$\{(\omega, x): x \in X_\alpha(\omega)\} \in \mathcal{A} \times \mathcal{B} \quad (2.1)$$

Para cada $\alpha \in [0; 1]$, donde $X_\alpha: \Omega \rightarrow \bar{\mathcal{F}}_0(\mathbb{R}^n)$ está definido por:

$$X_\alpha(\omega) = \{x \in \mathbb{R}^n: X(\omega, x) \geq \alpha\}. \quad (2.1)$$

Recalcando que $\mathcal{A} \subseteq \Omega$ y donde \mathcal{B} denota subconjuntos de Borel en \mathbb{R}^n . Los conjuntos de Borel para el ámbito difuso, junto a una mayoría de sus propiedades son representados de manera muy similar a la teoría de probabilidad (Willks, 1963). Se podría abarcar de manera muy amplia los aspectos estadísticos y por ende los matemáticos que dan un mayor soporte a las variables difusas, pero al igual que en la exposición de números difusos, nos inclinaremos por un punto en particular, el cual es los valores esperados difusos.

Expresado de esa manera son los conjuntos o mejor dicho los números difusos trapecoidales (es claro ver que un número triangular es un caso particular de un número trapecoidal) los cuales nos serán más fáciles de manejar.

¹⁶ La métrica sobre números difusos escapa al enfoque estadístico del avance, sin embargo se detalla la misma en anexos.

¹⁷ El concepto de conjunto compacto encaja perfectamente a la definición de función convexa, pero el primero es más general en topología una rama de matemáticos.

Considerando ahora una variable aleatoria difusa de tipo discreto $X: \Omega \rightarrow \bar{\mathcal{F}}_0(\mathbb{R})$ tal que $P(X = u_i) = p_i, i = 1, 2, \dots$, donde $u_i: \mathbb{R} \rightarrow [0; 1]$ tienen un soporte continuo y compacto. Es claro ver que con el uso de las operaciones aritméticas de suma y producto sobre el campo $K = \mathbb{R}$ expuestos en el apartado anterior, el término $u_1 + u_2 + \dots + u_i + \dots = \sum_{i=1}^{\infty} u_i$ también es un número difuso, y cualquier combinación lineal de esta secuencia $\{u_i\}$ es del mismo tipo.

Definición 2.10. (Valor Esperado) Para una f.r.v. discreta X tal que $P(X = u_i) = p_i$ y con $u_i: \mathbb{R} \rightarrow [0; 1]$ tienen un soporte continuo y compacto $\forall i \in \mathbb{N}$ el valor esperado de X viene dado por:

$$E[X] = \sum_{i=1}^{\infty} p_i * u_i \quad (2.1)$$

Nótese que no existe la condición de que sean números trapezoidales o triangulares, para el cálculo del valor esperado. En la Ilustración 7 se muestra un ejemplo de variable aleatoria discreta en la cual se puede aplicar el valor esperado. Existen otras definiciones sobre el variables aleatorias difusas (Nguyen, 1977), estas hacen énfasis en su valor esperado con el manejo bajo el principio de extensión y enfocado al uso variables lingüísticas.

Sea X una variable lingüística que toma valores en $U = (u_1, u_2, \dots, u_k)$ y $U \subseteq \mathbb{R}$, con probabilidades lingüísticas asignadas de la lista (A_1, A_2, \dots, A_k) . Para la esperanza lingüística de X (respecto a la lista) viene dado como la siguiente expresión:

$$E[X] = u_1 * A_1 + u_2 * A_2 + \dots + u_k * A_k \quad (2.1)$$

El significado de esta esperanza se basa en el principio de extensión de Zadeh. Pero al fijar los valores en cada conjunto difuso A_i en elementos x_i inicialmente se tiene la función de pertenencia la k-tupla de elementos x_i .

$$\mu_{A_1 \times \dots \times A_k}(x_1, \dots, x_k) = \bigwedge_{i=1}^k \mu_{A_i}(x_i)$$

Ahora bien se fijan los valores de los u_i en un espacio compuesto $S = \{u_i \in [0, 1]\}$ con $S \subset \mathbb{R}^k$, con la condición de $u_1 + \dots + u_k = 1$, quedando de esta manera la función de pertenencia de la esperanza $\tilde{E}[X]$ como:

$$\mu_{\tilde{E}[X]}(z) = \bigvee_{x_1 * u_1 + \dots + u_k * u_k = z}^{u_1 + \dots + u_k = 1} [\bigwedge_{i=1}^k \mu_{A_i}(x_i)]$$

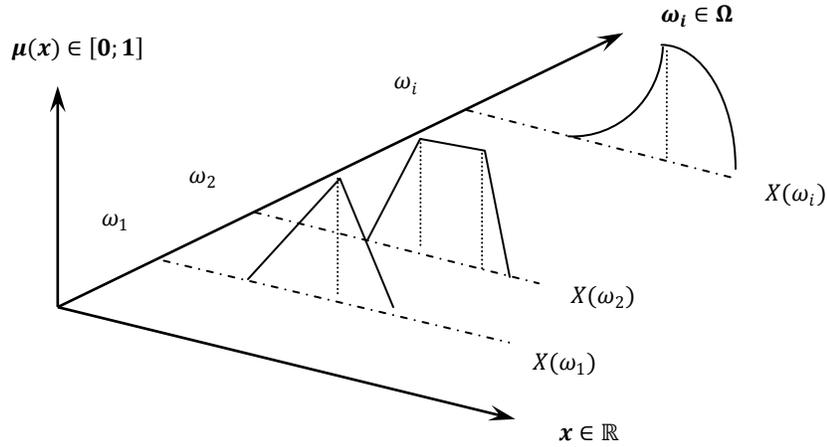


Ilustración 7. Representación de una variable difusa

Al particularizar el uso de la esperanza en el caso de los números trapezoidales se ha de considerar los siguientes:

- Cada número trapezoidal u_i se ha de manejar por la nueva cuarteta denotada como: $u_i = (l_i; m_i^l; m_i^r; r_i)$.
- La probabilidad asociada a cada número u_i es $p_i = P(u_i)$.

En el desarrollo de $E[X]$ se usa el hecho de la suma de dos NDT¹⁸ como $\tilde{U} + \tilde{W}$ y el hecho de la multiplicación por un escalar $k > 0$, resulta en $\tilde{U} = (ku_1, ku_2, ku_3, ku_4)$

Demostración:

$$\begin{aligned}
 E[X] &= \sum_{i=1}^{\infty} p_i * u_i \\
 &= \sum_{i=1}^{\infty} p_i * (l_i; m_i^l; m_i^r; r_i) \\
 &= \sum_{i=1}^{\infty} (p_i * l_i; p_i * m_i^l; p_i * m_i^r; p_i * r_i) \\
 &= \left(\sum_{i=1}^{\infty} p_i * l_i; \sum_{i=1}^{\infty} p_i * m_i^l; \sum_{i=1}^{\infty} p_i * m_i^r; \sum_{i=1}^{\infty} p_i * r_i \right)
 \end{aligned}$$

Con este hecho se verifica que $E[X]$ también es un número difuso trapezoidal, y para el cálculo de la varianza $V[X]$ en este tipo particular de variable, se necesita el concepto de distancia entre números difusos, ya que la varianza representa una discrepancia entre cada valor de la variable y su esperanza $E[X] = \tilde{M}$.

Al definir una distancia entre dos números difusos: $d(\tilde{U}, \tilde{W})$ como la distancia entre \tilde{U} y \tilde{W} , por otro lado al concepto de distancia se introduce los ya conocidos α – cortes

¹⁸ Se usará como abreviación de números difusos trapezoidales.

mediante $d(U_\alpha, W_\alpha)$ y otro concepto de distancia es el de distancia entre los centroides¹⁹ de \tilde{U} y \tilde{W} denotados por m_U y m_W dicha distancia se denota $d(m_U, m_W)$.

Definición 2.11. (Varianza de una FRV) Para una f.r.v. discreta X tal que $P(X = u_i) = p_i$ y con $u_i: \mathbb{R} \rightarrow [0; 1]$ tienen un soporte continuo y compacto $\forall i \in \mathbb{N}$, con esperanza $E[X] = \tilde{M}$ la varianza de X se define como:

$$V[X] = \sum_{i=1}^{\infty} d(u_i, \tilde{M}) * p_i \quad (2.1)$$

La elección de la función numérica no se pondrá en discusión, ya que existen muchas variantes en su manejo y justificación. Sin embargo es necesario aclarar que dicha distancia denotada $d(\tilde{U}; \tilde{W}) \in \mathbb{R}^+$ con $\tilde{U}; \tilde{W} \in \tilde{\mathbb{R}}$ es denominada distancia como tal si y solo si:

- $d(\tilde{U}, \tilde{W}) \geq 0 \quad \forall \tilde{U}, \tilde{W} \in \tilde{\mathbb{R}}$
- Si $\tilde{U} = \tilde{W} \quad \Rightarrow \quad d(\tilde{U}, \tilde{W}) = 0$
- $d(\tilde{U}, \tilde{W}) = d(\tilde{W}, \tilde{U}) \quad \forall \tilde{U}, \tilde{W} \in \tilde{\mathbb{R}}$
- $d(\tilde{U}, \tilde{Z}) \leq d(\tilde{U}, \tilde{W}) + d(\tilde{W}, \tilde{Z}) \quad \forall \tilde{U}, \tilde{W}, \tilde{Z} \in \tilde{\mathbb{R}}$

2.1.5. Métodos de Defuzzificación

En la mayoría de los métodos que se usa en campo de la teoría difusa, se obtienen resultados en un sentido difuso, pero para la presentación de dichos resultados casi siempre se usa valores en el eje real con esto se hace uso de los denominados “Métodos de Defuzzificación”.

Los métodos citados lo que hacen en sí, es un transformación de un conjunto difuso (generalmente un número difuso) al eje real, representando de esta manera los resultados obtenidos por número real.

Se presentan a continuación los métodos de mayor uso y difusión dentro del proceso de Defuzzificación de elementos principales o conjuntos:

Máxima Pertenencia (Método Height)

Este método consiste básicamente en determinar un valor " x_0 ", el cual cumpla lo siguiente condición sobre su función de pertenencia " μ_A " para un conjunto difuso A :

$$\mu_A(x_0) \geq \mu_A(x) \quad \forall x \in \mathcal{X} \quad (2.14)$$

Claramente este método se basa en la determinación de un término similar a la moda, que se usa en frecuentemente en temas de variables aleatorias.

¹⁹ Los métodos de defuzzificación darán un detalle más amplio de los centroides.

Centro de Gravedad (Método del centroide)

Básicamente este método calcula un similar o equivalente a lo que se conoce como valor esperado $E[X]$ en estadística. Aquí es denominado por " x_0 " el cual se determina mediante la siguiente expresión:

$$x_0 = \frac{\int_{-\infty}^{+\infty} x * \mu_A(x) dx}{\int_{-\infty}^{+\infty} \mu_A(x) dx} \quad (2.15)$$

Para casos discretos de " μ_A " se debe de sustituir la integral por una sumatoria respectiva, teniendo el mismo significado y valor.

Promedio ponderado

Restringido a solo funciones de pertenencia simétricas, este es el método más usado, pero se maneja sobre varios conjuntos A_i , y los valores " x_i " denotan los centroides para cada conjunto difuso A_i , este valor se calcula mediante la siguiente expresión:

$$x_0 = \frac{\sum x_i * \mu_{A_i}(x_i)}{\sum \mu_{A_i}(x_i)} \quad (2.16)$$

Promedio de Máximos (Middle of maxima)

Usado principalmente en funciones trapezoidales y pseudo-trapezoidales²⁰, este método encuentra dos puntos denotados por x_a y x_b , que cumplen que $\mu_A(x_a) = \mu_A(x_b) = 1$, además que para todo $x \in [x_a; x_b]$ se verifica que $\mu_A(x) = 1$. Se calcula el valor defuzzificado " x_0 " mediante la siguiente ecuación:

$$x_0 = \frac{x_a + x_b}{2} \quad (2.17)$$

La elección del método de defuzzificación es sin duda, una elección una tanto subjetiva, del experto en si, además que se deben de tomar aspectos muy específicos en cada caso de aplicación.

2.1.6. Conjuntos difusos sobre Variables Lingüísticas

Como se mencionó antes, la aplicación de los conjuntos difusos es amplia, se hará énfasis ahora en un tipo aplicación muy importante este es el de variables lingüísticas.

El concepto de número difuso juega un papel fundamental al formular el manejo de variables cuantitativas difusas. Cuando adicionalmente los números difusos representan conceptos lingüísticos, tales como: muy pequeño, pequeño, mediano y otros más, son interpretados en un contexto particular, los resultados construidos son denominados usualmente como *variables lingüísticas*.

²⁰ Este tipo de función es una modificación a la función trapezoidal, y también define a un conjunto difuso.

Cada estado de la variable lingüística es expresado por términos lingüísticos, y estos serán interpretados como números difusos, definidos en términos de la variable base. Una variable base es una variable en el sentido clásico, ejemplificado por cualquier variable física (ejemplo: temperatura, presión, voltaje, humedad, etc.) así como también cualquier otra variable numérica (edad, tasa de interés, rendimiento, salario etc.). En una variable lingüística, los términos lingüísticos representan valores aproximados de la variable base, relacionados a alguna aplicación, y de esta manera son capturados por números difusos.

Tal como se mencionó en el capítulo anterior, cada variable lingüística es completamente caracterizada por la quintupla (v, T, X, g, m) en el que “ v ” es el nombre de la variable, “ T ” es el conjunto de términos lingüísticos de “ v ” que se refieren a la variable base cuyo rango de valores es el conjunto universal “ X ”, “ g ” es una regla sintáctica (una gramática) para generar términos lingüísticos, y “ m ” es una regla semántica que asigna a cada término lingüístico “ $t \in T$ ”.

Un ejemplo de variable lingüística se representaría el rendimiento de una entidad u objeto (persona, máquina, organización, método, etc.) en este contexto los cinco términos lingüísticos son: *muy pequeño*, *pequeño*, *medio*, *grande* y *muy grande*. Cada uno de los términos lingüísticos básicos se le asigna uno de los cinco números borrosos por una regla semántica. Los números borrosos, cuyas funciones de pertenencia son habitualmente formas trapezoidales.

2.1.7. Series de Tiempo Difusas

Antes de exponer de manera formal los conceptos de Serie de Tiempo Difusa, y con fines netamente comparativos se ve necesario citar a Granger y Newbold (1986) quienes describen a las series de tiempo como: “*una secuencia de observaciones ordenadas por un parámetro de tiempo*” las series de tiempo se pueden medir de manera continua o discreta.

Las series temporales se pueden registrar instantáneamente y de manera constante. La mayoría de las mediciones en las ciencias ya sean sociales, económicas o de otra índole, se hacen en intervalos regulares o irregulares, y estos datos de series de tiempo son denominados discretos o continuos correspondientemente. Acumulaciones de lluvia medidos discretamente a intervalos regulares sería un ejemplo. Otros pueden ser reunidos a partir de las observaciones individuales para hacer un resumen estadística, medida a intervalos regulares a lo largo del tiempo.

En el tratamiento de las series temporales, se tienen desarrolladas distintas técnicas y modelos a lo largo del avance del tema, pero tal como se mencionó en un inicio, la naturaleza de las variables y la crítica en la medición de los datos, trazan un nuevo campo para el manejo de datos en un contexto *borroso o difuso*.

Metodología de Box & Jenkins (Modelo ARIMA)

En el análisis de series de tiempo es de entender que en la actualidad la técnica más usada es la de modelos ARIMA, este misma fue desarrollada por *Box & Jenkins* para la década de los 70. La metodología se basa en el proceso de estimación, para la siguiente ecuación de un determinado modelo ARIMA ($p; d; q$) :

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2.18)$$

En el cual la serie de datos de " W_t " se define en función de la serie original " Y_t " mediante la siguiente relación: $W_t = \nabla^d Y_t$, en donde el operador diferencia " ∇ " se define como el cambio absoluto entre un determinado periodo y el anterior es decir: $\nabla Y_t = Y_t - Y_{t-1}$, y similarmente se generaliza para el caso " ∇^d ".

En (2.18) los términos $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ son los denominados ruidos blancos, que poseen distribución normal con media nula y varianza constante. Como es de entender en un modelo probabilístico, se han de plantear supuestos sobre parámetros, distribución, restricciones, y no es la excepción los modelos ARIMA.

Metodología de Song & Chissom's (Serie de Tiempo Difusa)

En el progreso de la teoría difusa para diversas áreas, se han hecho grandes avances, uno de ellos en particular son las *Series de Tiempo Fuzzy*, estas manejan series temporales, tal como se señaló en el capítulo anterior, fueron desarrollada por Song & Chissom's, y se vinculan los cambios de los datos temporales mediante relaciones de conjuntos difusos:

Definición 2.12. (Serie de tiempo difusa) Sea $Y(t)$ con $(t = 0, 1, 2, \dots)$, un subconjunto de \mathbb{R} , siendo el universo de discurso \mathcal{X} , para los conjuntos difusos $\mathfrak{f}_i(t): i = 1, 2, 3, \dots$. Si $F(t)$ es la colección de $\mathfrak{f}_i(t)$, entonces $F(t)$ se denomina Serie de Tiempo Difusa bre $Y(t)$.

Definición 2.13. (Left & Right) Dados los conjuntos difusos A_i y A_j , que se representan mediante $A_i = F(t - 1)$ y $A_j = F(t)$, una relación lógica difusa puede definirse como $A_i \rightarrow A_j$, donde A_i y A_j son denominados el lado izquierdo (LHS) y el lado derecho (RHS) de la relación lógica difusa, respectivamente.

En estas definiciones se ha planteado como espacio referencial \mathcal{X} de la serie temporal original, a lo largo del avance varios autores han diseñado y desarrollado metodologías en base los cambios, como son los absolutos, relativos, transformaciones etc. de un periodo con respecto al otro. Esta reciprocidad de cambio se especifica mediante el siguiente esclarecimiento.

Definición 2.14. (Relación difusa) Si existe una relación denotada por $R(t; t - 1)$, de tal manera que $F(t) = F(t - 1) \otimes R(t; t - 1)$ donde \otimes representa una operación, entonces $F(t)$ se dice que es causado por $F(t - 1)$. La relación entre $F(t)$ y $F(t - 1)$ puede representarse como:

$$F(t - 1) \rightsquigarrow F(t) \quad (2.19)$$

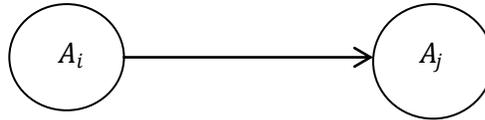


Ilustración 8. Relación difusa de 1er orden

Los aspectos más importantes sobre esta relación "R" se han de detallar en posteriores secciones, también es claro ver que esta relación es de primer orden ya que depende de un plazo atrás, generalizandose para más rezagos se tiene lo siguiente:

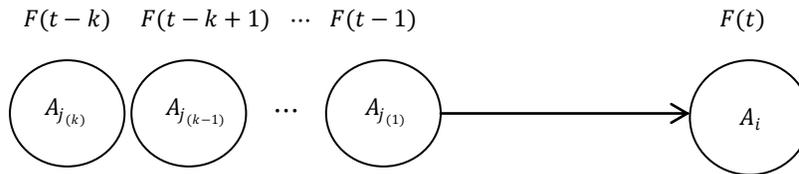


Ilustración 9. Relación difusa de orden k

Definición 2.15. (Relación de orden k) Sea $F(t)$ una serie de tiempo difusa. Se dice que $F(t)$ es causado por $F(t - 1), F(t - 2), \dots, F(t - k)$, entonces la relación difusa de orden k , está representada por:

$$F(t - 1), F(t - 2), \dots, F(t - k) \rightsquigarrow F(t) \quad (2.20)$$

Introducida originalmente por Chen, esta definición hace referencia a modelos de orden mayor, cuando se las maneja. El aporte más importante de Chen es la siguiente tesis, sobre las relaciones difusas dentro de una serie temporal.

Definición 2.16. (Serie invariante) Suponga que $F(t)$ es causado solamente por $F(t - 1)$, mediante la relación $F(t) = F(t - 1) \times R(t; t - 1)$. Si para todo "t" el término R es independiente de "t" entonces $F(t)$ se denomina serie de tiempo difusa-invariante, de lo contrario se considerara serie de tiempo difusa-variante.

De manera algo similar a la relación de orden k , se tiene las relaciones de grupo difusas, las cuales relacionan una misma de un conjunto difuso, el cual causa efectos en distintos conjuntos de la serie.

Definición 2.17. (Grupo de relación difuso) Sea que $A_i, A_{j_1}, A_{j_2}, \dots, A_{j_m}$ conjuntos difusos sobre la serie de tiempo fuzzy $F(t)$, Si $A_i \rightarrow A_{j_1}; A_i \rightarrow A_{j_2}; \dots; A_i \rightarrow A_{j_m}$, entonces se puede establecer una relación de grupo mediante: $A_i \rightarrow A_{j_1}; A_{j_2}; \dots; A_{j_m}$.

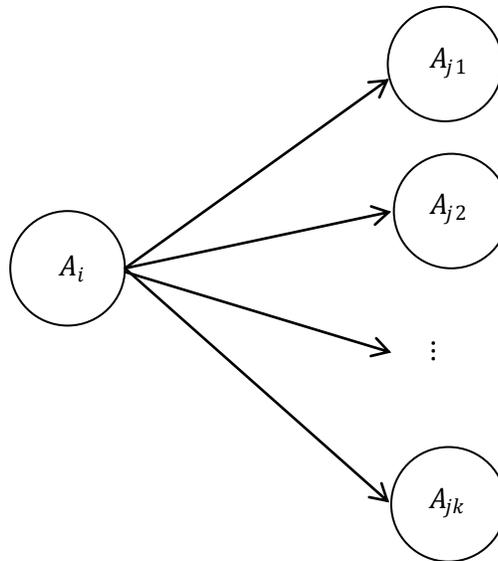


Ilustración 10. Relación difusa de grupo

Las relaciones con el mismo conjunto difuso sobre el lado izquierdo se pueden clasificar en una relación de grupo difuso, también se les conoce como grupos difusos de relaciones lógicas o *FLRG* (por sus siglas en inglés).

2.2. Modelos de series de tiempo difusas

Desde el trabajo de Zade, se desplegaron métodos heterogéneos de análisis de información bajo el concepto de datos difusos, estos mismos posteriormente aportaron al adeudo guía sobre Series de Tiempo Fuzzy propuesto. Entre estos aportes se han plasmado modelos análogos al de Song & Chissom's y sobre todo en base al enfoque de Chen.

Considerando los aportes de los dos autores antes mencionados, un gran colectivo de estos modelos, comparten entre si las siguientes etapas en la construcción del modelo en si:

2.2.1. Etapa 1ra: Creación de Clusters & Relaciones Difusas

En esta etapa se decide como se va a dividir el espacio de discusión, o como en el nombre lo indica se decide la creación los clusters. La más comúnmente usada, divide el

recorrido de la variable en intervalos de amplitud constante, una mayoría de autores de los modelos usa esta opción.

Una segunda alternativa fue introducida (Arutchelvan, Sivasta, & Jaganathan, 2010), estableciéndose como una re-división de los clusters, la cual estaba sujeta al número de elementos que estos contenían, dando lugar implícitamente al uso de cuantiles²¹ dentro de todo el proceso.

Cabe aclarar el manejo de *datos atípicos* en esta etapa es muy importante, ya que estos inciden en la formación de clusters. El método el cual aborda esto, es la corrección que Rúni propone para la detección de datos atípicos (Runi, 2006). En esta detalla la el tratamiento y detección de los mismos mediante el uso de la desigualdad de Chevisef, y el cálculo de algunas estadísticas, dicho de una manera más sencilla se usa la estandarización de los datos, pero con el concepto de distancia media y desviación estándar, como se ve a continuación:

$$AD = \frac{1}{n-1} \sum_{i=1}^{n-1} [x_{(i+1)} - x_{(i)}] \quad (2.21)$$

$$\sigma_{AD} = \sqrt{\frac{1}{n} \sum_{i=1}^n (|x_i| - AD)^2} \quad (2.22)$$

Para la primera expresión se debe aclarar que los datos originales $\{x_i\}$ deben de ordenarse ascendentemente en un conjunto denotado por: $\{x_{(i)}\}$ para luego trabajar sobre ese nuevo conjunto.

El numero de Clusters a ser manejados es muy heterogéneo entre los autores, sin embargo Chen manejo siete conjuntos en un inicio, y el hecho de que se recomienda un uso de número impar de clusters, se respalda en el habituado manejo de test que aplican términos lingüísticos como son las escalas de Likert²² y otras más. Desde el punto de vista estadístico, se sabe que es conveniente agrupar los datos en \sqrt{n} grupos, o también siguiendo la regla de Stuges $\ln(n) + 1$.

2.2.2. Etapa 2da: Determinación de relaciones

Una vez creados los clusters y la discretización de los datos, se han de establecer las relaciones difusas entre los clusters o conjuntos creados. Pero como se verá a continuación, cada modelo e incluso del mismo autor, diverge demasiado como para poner una presentación más general, así que se mencionara a los modelos de mayor relevancia, en términos explícitos.

²¹ Claros ejemplos de los cuantiles son: cuartiles, deciles, percentiles, etc.

²² La escala de likert es usada en una gran variedad de test psicológicos, en donde se tiene 5 niveles: muy en desacuerdo (MD), en desacuerdo (D), indeciso (I), de acuerdo (A) y muy de acuerdo (MA) respecto a una afirmación o posición mencionada en cada pregunta.

Modelo de Chen (1996)

Fase 1: Definir el universo de discurso e intervalos para reglas de abstracción. Basándose en el dominio de problema, el universo de discurso se puede definir como: $U = [Inicio; Final]$. La longitud de los intervalos ha determinarse es constante para todos.

Fase 2: Definir conjuntos difusos basados en el universo de discurso y datos históricos.

Fase 3: Fuzzificar reglas observadas.

Fase 4: Establecer las relaciones difusas (FLRs) y de grupo (FLRG) ellos sobre la base de los estados actuales de los datos de relaciones lógicas difusas.

Fase 5: Pronosticar Sea: $F(t - 1) = A_i$ para los siguientes casos:

- **1er Caso:** Si la relación lógica difusa de A_i es vacía, es decir: $A_i \rightarrow \emptyset$; entonces el valor pronosticado de $F(t)$ es A_i .
- **2do Caso:** Si sólo hay una relación lógica difusa en secuencia: $A_i \rightarrow A_j$, entonces el valor pronosticado de $F(t)$ es igual a A_j .
- **3er Caso:** Si hay una relación lógica difusa en secuencia: $A_i \rightarrow A_{j1}, A_{j2}, \dots, A_{jk}$, entonces el valor pronosticado de $F(t)$ es igual a el conjunto $A_{j1}, A_{j2}, \dots, A_{jk}$.

Fase 6: Defuzzificar, si el valor pronosticado está comprendido por: $A_{j1}, A_{j2}, \dots, A_{jk}$, mediante un método de agregación.

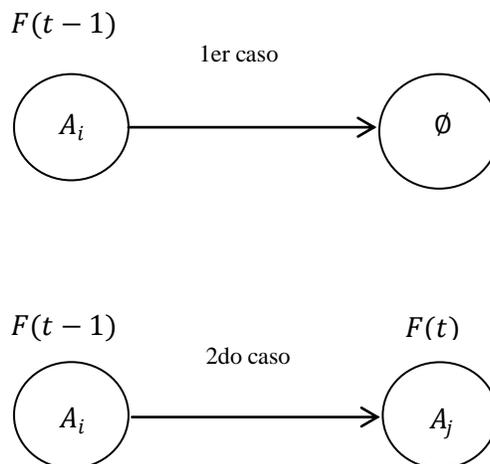


Ilustración 11. Primer y segundo caso en la regla de Chen

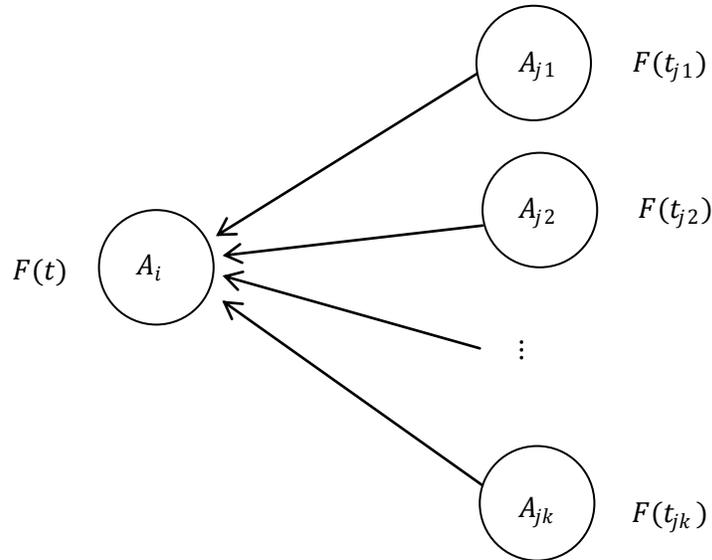


Ilustración 12. Tercer caso en la regla de Chen

Modelo de Yu's (2005)

Fase 1: Definir el universo de discurso y sub-intervalos. Basados sobre los valores mínimos y máximos del conjunto de datos, D_{min} y D_{max} son definidas. A continuación, elija dos números arbitrarios positivos que son D_1 y D_2 , esto con el fin de dividir en intervalos uniformemente denotados por: $U = [D_{min} - D_1 ; D_{max} + D_2]$.

Fase 2: Definir conjuntos difusos basados en el universo de discurso y datos históricos.

Fase 3: Fuzzificar reglas observadas.

Fase 4: Establecer las relaciones lógicas difusas (revisar el método de Chen's). Los FLR's recurrentes ahora se consideran: $A_i \rightarrow A_i$, además de contemplar la frecuencia de los mismos. Por ejemplo para cuatro relaciones: $A_1 \rightarrow A_3, A_1 \rightarrow A_1, A_1 \rightarrow A_5, A_1 \rightarrow A_1$, entonces la relación en lógica difusa del grupo está dada por: $A_1 \rightarrow A_3, A_1, A_5, A_1$.

Fase 5: Ídem al algoritmo de Chen.

Fase 6: Defuzzificar. Sea el pronóstico: $A_{j_1}, A_{j_2}, \dots, A_{j_k}$ para el valor de $F(t)$. La matriz de defuzzificación es $M(t) = [m_{j_1}, m_{j_2}, \dots, m_{j_k}]$ donde m_{j_l} es el punto medio de A_{j_l} respectivo.

Fase 7: Determinar los ponderadores de $A_{j_1}, A_{j_2}, \dots, A_{j_k}$ que serán calculados por la siguiente matriz: $W(t) = [w_1, w_2, \dots, w_k]$ donde $w_i = i / \sum i$; con $i = 1, 2, \dots, k$.

Fase 8: Calcular el valor del pronóstico para cada $F(t)$, el cual esta expresado por:
 $F(t) = M(t) \times W(t)^t$.

Modelo de Lee's (2008)

Fase 1: Ídem a Yu's

Fase 2: Establecer las relaciones de conjuntos difusos (valores lingüísticos) para cada observación en el conjunto de datos.

Fase 3: Establecer las relaciones difusas.

Fase 4: Establecer las relaciones difusas para todos los grupos de FLR's.

Fase 5: Seleccionar el mejor orden de FLR's. Calculando la matriz de fluctuaciones.

Fase 6: Pronosticar

Fase 7: Defuzzificar, usando la regla de Yu's.

Fase 8: Si el pronóstico de $F(t)$ es $A_{j1}, A_{j2}, \dots, A_{jk}$. La matriz de ponderadores se obtendrá como: $W(t) = (1/\sum w_h) * [w_1, w_2, \dots, w_k]$ con $w_h = c^{h-1}$; $h = 1, 2, \dots, k$.

Fase 9: Calcular el valor pronosticado del término $F(t)$ mediante: $F(t) = M(t) \times W(t)^t$ con $M(t) = [m_{j1}, m_{j2}, \dots, m_{jk}]$ donde m_{jl} es el punto medio de cada A_{jl} respectivo.

2.2.3. Etapa 3ra Calculo de Pronósticos

Tras la conformación de cada clúster y el establecimiento de las relaciones entre estos, ahora se procede al cálculo de los pronósticos. Como se vio anteriormente al final de cada modelo se procede al cálculo mediante el uso de ponderadores, este concepto que si bien parece sencillo, se pondrá a discusión de manera más profunda en el apartado (2.3) ya que se expondrán elementos necesarios en dicho apartado.

2.2.4. Críticas y Avances en las STD

Desde su inicio se tilda a la teoría de datos difusos, como algo pernicioso, como una manera más de la estadística. Las ventajas que ya se menciono fueron para el análisis de series de tiempo, siendo condiciones tales como: la normalidad, estacionariedad los cuales no son necesarios, puede trabajar incluso algunos de los datos no están disponibles, logra ocuparse en la serie de tiempo de manera lingüística, facilitando de esta manera una mejor interpretación de resultados.

No se puede citar todos los avances de esta metodología ya que estos mismos siguen proponiendo nuevas técnicas, sin embargo a continuación se presentara algunas de las incorporaciones mas enfatizadas con el área de estadística:

- Análisis de tendencias y estacionareidad , Bernd Möller (2006).
- Incorporación de modelos multivariantes, Bulut & Shigeru (2012).
- Modelos ARIMA, denominados FARIMA por presentar una componente difusa, Tseng & Tzeng (2001), sin embargo esta metodología encaja más a el área denominada como regresión difusa.

Aclarando que existen otros aportes y avances más entre estos modelos, son los antes mencionados los más sobresalientes desde un punto de vista de análisis estadístico de series de tiempo.

2.2.5. Medidas de Bondad de ajuste

El proceso de evaluar los resultados obtenidos a partir de estos modelos, recae en medidas generalmente usadas por los autores, los cuales recomiendan el uso de las siguientes medidas de bondad de ajuste:

Error cuadrático medio (MSE): Siendo el más usado en entre estos modelos de datos difusos y otros que no lo son, esta medida de bondad de ajuste se define como:

$$MSE = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2 \quad (2.23)$$

Error medio absoluto (MAE): Otros sugieren el manejo predominante de esta medida para hacer sus comparaciones (Runi, 2006).

$$MAE = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t| \quad (2.24)$$

Error medio porcentual (MAPE): Diversos autores basan su trabajo sobre las variaciones relativas y manejan este indicador como referencia de comparaciones (Arutchelvan, Sivasta, & Jaganathan, 2010).

$$MAPE = \frac{100}{T} \sum_{t=1}^T \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (2.25)$$

Todas estas medidas de bondad de ajuste también se manejan en otras metodologías de pronósticos que no son las de datos difusos. Por lo cual nos sirven como indicador de comparación con estas otras técnicas, como los modelos ARIMA y los métodos de suavización.

2.3. Medias generalizadas

Al parecer cuando se nombra el término de media o lo que comúnmente se conoce como promedio, nos da a entender definiciones bastante elementales. Sin embargo la teoría detrás de estos estadígrafos comúnmente usados, es bastante extensa y con un respaldo muy fuerte.

Las medias generalizadas se desarrollaron por Hardy & Littlewood en 1934 fundamentándose estos en publicaciones de O. Hölder (1889), posteriormente se puso un mayor énfasis en su aplicación gracias a la base que se puso en las “funciones de ingreso igualmente distribuidas” introducidas por Anthony B. Atkinson en la década de los 70. Ahí se expone que cada función es un estándar de ingreso que da mayor peso a los ingresos de los pobres, sin ignorar los ingresos de los “menos” pobres, o dicho de otro modo, se otorga una menor importancia o peso a ingresos altos (Atkinson, 1970).



Ilustración 13. Anthony B. Atkinson autor de:
“On the Measurement of Inequality”

2.3.1. La F-media Generalizada

Los promedios más usados en la mayoría de aplicaciones son: la media aritmética, media armónica y la media geométrica, estas últimas también denominadas medias pitagóricas, pertenecen a un sub grupo de operadores conocidos como medias potencia o medias ordinarias.

Definición 2.18. (Hardy & Littlewood) Sea un conjunto de n números no negativos denotados por a_1, a_2, \dots, a_n ($a_i \geq 0$) y sea un número real " r ", se denota la **Media ordinaria de orden r** , mediante la siguiente expresión:

$$\mathfrak{R}_r = \left(\frac{1}{n} \sum_{i=1}^n a_i^r \right)^{1/r} \quad (2.26)$$

Si bien se generan distintos estadígrafos como son: media cuadrática, media cubica y otros más, el proceso de verificación se centra en las 3 pitagóricas mencionadas en un inicio, el mismo es el siguiente:

Media aritmética: Es inmediato al reemplazar $r = 1$ en la ecuación (2.26) obteniéndose:

$$\mathfrak{R}_r = \frac{1}{n} * \sum_{i=1}^n a_i$$

Media geométrica: Definamos primero a: $\omega_r = \ln(\mathfrak{R}_r) = \frac{1}{r} * \ln(\sum_{i=1}^n \frac{1}{n} * a_i^r)$ ahora bien:

$$\begin{aligned} \lim_{r \rightarrow 0} \omega_r &= \lim_{r \rightarrow 0} \left[\frac{1}{r} * \ln(\sum_{i=1}^n \frac{1}{n} * a_i^r) \right] \\ &= \lim_{r \rightarrow 0} \frac{\ln(\sum_{i=1}^n \frac{1}{n} * a_i^r)}{r} \end{aligned}$$

Esta última expresión de límite al evaluar en 0, nos da “0” tanto en numerador como en el denominador, así aplicando regla de L’Hopital’s, se obtiene:

$$\begin{aligned} \lim_{r \rightarrow 0} \omega_r &= \lim_{r \rightarrow 0} \frac{[\ln(\sum_{i=1}^n \frac{1}{n} * a_i^r)]'}{1} \\ &= \lim_{r \rightarrow 0} \frac{[\sum_{i=1}^n \frac{1}{n} * a_i^r]'}{\sum_{i=1}^n \frac{1}{n} * a_i^r} \\ &= \lim_{r \rightarrow 0} \frac{1}{\sum_{i=1}^n \frac{1}{n} * a_i^r} * [\sum_{i=1}^n \frac{1}{n} * e^{r * \ln(a_i)}]' \\ &= \lim_{r \rightarrow 0} \frac{1}{\sum_{i=1}^n \frac{1}{n} * a_i^r} * \sum_{i=1}^n \frac{1}{n} * \ln(a_i) * e^{r * \ln(a_i)} \\ &= \frac{1}{\sum_{i=1}^n \frac{1}{n} * a_i^0} * \sum_{i=1}^n \frac{1}{n} * \ln(a_i) * e^{0 * \ln(a_i)} \\ &= \frac{1}{\sum_{i=1}^n \frac{1}{n}} * \sum_{i=1}^n \frac{1}{n} * \ln(a_i) \\ &= \sum_{i=1}^n \frac{1}{n} * \ln(a_i) \quad \therefore \end{aligned}$$

Donde este último término es el igual a: $\omega_0 = \ln \left[(\prod a_i)^{\frac{1}{n}} \right]$ y el cual nos lleva a:

$$\begin{aligned} \mathfrak{R}_0 &= \exp(\omega_0) \\ &= \exp \left\{ \ln \left[(\prod a_i)^{\frac{1}{n}} \right] \right\} \\ \mathfrak{R}_0 &= (\prod a_i)^{\frac{1}{n}} \quad \therefore \end{aligned}$$

Media armónica: La verificación de este valor, también es inmediata al reemplazar $r = -1$ en la ecuación (2.26), obteniéndose:

$$\mathfrak{R}_r = n / [\sum_{i=1}^n (1/a_i)]$$

Propiedades de la F-media generalizada

Como toda función, las medias ordinarias o generalizadas poseen ciertas propiedades, citaremos las mas resaltantes mencionadas por Hardy & Littlewood. Una gran mayoría de estos promedios son considerados como “funciones homogéneas” esto es que:

$f(a * x) = a * f(x)$ para algún valor constante "a". Por otro la más importante propiedad es la que a continuación se citara:

Desigualdad de Medias: Sean p y q dos números positivos tales que $p < q$, se cumple para un conjunto de datos a_1, a_2, \dots, a_n ($a_i \geq 0$) que:

$$\mathfrak{R}_p = \left(\frac{1}{n} \sum_{i=1}^n a_i^p \right)^{1/p} \leq \mathfrak{R}_q = \left(\frac{1}{n} \sum_{i=1}^n a_i^q \right)^{1/q} \quad (2.27)$$

Prueba: Antes se debe de tener en cuenta la *desigualdad de Jensen*²³, la misma afirma que si se tiene una función convexa $f(x)$, y se disponen de constates $\lambda_1, \lambda_2, \dots, \lambda_n$ tales que $\sum \lambda_i = 1$, se verifica la siguiente inecuación:

$$f\left(\sum_{i=1}^n \lambda_i * x_i\right) \leq \sum_{i=1}^n \lambda_i * f(x_i) \quad (2.28)$$

Como $p < q$, el término $\frac{q}{p} - 1 > 0$ asegurando la creación de una función: $f(x) = x^{\frac{q}{p}}$, para la cual se puede verificar el hecho de su convexidad. Tomando $\lambda_i = \frac{1}{n}$ verifica una suma igual a uno, además usamos esta nueva función en la expresión anterior, y posteriormente tomando raíz de orden q obtenemos:

$$\begin{aligned} f\left(\sum_{i=1}^n \lambda_i * x_i^p\right) &\leq \sum_{i=1}^n \lambda_i * f(x_i^p) \\ \left[\sum_{i=1}^n \lambda_i * x_i^p\right]^{\frac{q}{p}} &\leq \sum_{i=1}^n \lambda_i * (x_i^p)^{\frac{q}{p}} \\ \left[\sum_{i=1}^n \lambda_i * x_i^p\right]^{\frac{1}{p}} &\leq \left[\sum_{i=1}^n \lambda_i * x_i^q\right]^{\frac{1}{q}} \\ \mathfrak{R}_p &\leq \mathfrak{R}_q \quad \therefore \end{aligned}$$

Es decir que a medida que aumentamos el valor del parámetro "r" en el conjunto de medias ordinarias \mathfrak{R}_r se obtendrá medias o promedios de mayor envergadura.

2.3.2. Medias Ponderadas

El concepto de ponderación dentro de los operadores promedio se abarca de manera más sencilla aun, ya que el término " $\frac{1}{n}$ " se reemplaza por: " w_i " relacionándolo para cada elemento a_i , donde estos son denominados ahora ponderadores del i-esimo elemento y verifican: $w_i > 0$ y $\sum w_i = 1$.

Para no caer de manera redundante en la verificación, se presenta la adaptación con la incorporación de estos ponderadores, en cada caso con: $\mathfrak{R}_r = \left(\sum_{i=1}^n w_i a_i^r\right)^{1/r}$, como se verá en la TABLA 1.

²³ La demostración de esta desigualdad escapa al avance del capítulo, pero se detalla en Anexos.

Los otros promedios ponderados generalizados son consecuentes a estos, pero se aclarara si es necesario su uso conforme se dé avance el tema de aplicación de los mismos.

TABLA 1. CLASIFICACION DE MEDIAS ORDINARIAS

Denominación	Notación	Ecuación
<i>Media aritmética ponderada</i>	\mathfrak{R}_1	$\sum w_i a_i$
<i>Media geométrica ponderada</i>	\mathfrak{R}_0	$\prod a_i^{w_i}$
<i>Media armónica ponderada</i>	\mathfrak{R}_{-1}	$1/\sum(w_i/a_i)$

2.3.3. Medias Ordenadas Ponderadas (OWA)

Tal como se ve sus siglas en ingles: *Ordered Weighted Averaging*, tratan de promedios que se calculan con un previo orden en los datos, en estadística estas funciones se conoce como *estadígrafos de posición*, siendo estos últimos un caso particular de los operadores OWA.

El término OWA fue introducido por Ronald R. Yager en 1988, proporcionando un método de puntuación el cual abarca un operador de comportamiento conjuntivo y disyuntivo, inclinando su aplicación a la elección de criterios de selección con el uso de la siguiente ecuación:

$$OWA_r = \left(\sum_{i=1}^n w_i a_{(i)}^r \right)^{1/r} \quad (2.29)$$

Para esta expresión la única diferencia es que el conjunto de datos $\{a_i\}$ deben de ordenarse ascendentemente en un nuevo conjunto de elementos denotados por: $a_{(i)}$ los cuales verifican que: $a_{(1)} \leq a_{(2)} \leq \dots a_{(n)}$, aclarando el hecho de elección para los ponderadores w_i debe ser adecuada a la aplicación a realizarse.

2.3.4. Aplicación de la F-media en las Series de Tiempo Difusas

Varias alternativas a la metodología de Chen se han manejado, la mayoría se apega al uso de promedios y todo esto incurre dentro del proceso para cálculo de pronósticos. A continuación se mencionara tipos de promedios usados por cada autor que los manejaron.

Es el autor principal **Chen (1996)** quien usa la media aritmética para el cálculo del valor pronosticado mediante:

$$\hat{F}_i(t) = \frac{1}{k} \sum_{h=1}^k m_{j,h} \quad (2.30)$$

Donde " $m_{j,h}$ " son los valores defuzzificados de los k elementos que están en *LHS* en la relación difusa, y para un conjunto Y $\hat{F}_i(t)$ es el valor pronosticado del elemento A_j que está en *LRS*.

Jilani & Burney (2008) proponen el uso de la media armónica ponderada en una infinidad de modelos (Jilani, Burney, & Ardil, 2008), con algunas diferencias sutiles, pero siempre bajo el mismo concepto general de manejo de promedios expresados por:

$$\hat{F}_i(t) = \frac{\sum_j w_j}{\sum_j \frac{w_j}{m_j}} \quad (2.31)$$

Donde los ponderadores w_j son evaluados de manera escalonada es decir: $w_j = 0 ; 0.5 ; 1$, dentro de cada conjunto de elementos y los términos m_j son los valores Defuzzificados asociados a estos. Esta técnica no usa las reglas de Chen y es por eso que no se menciono en un inicio.

Lee's (2008): En su algoritmo más concretamente en la fase del cálculo, usando ponderadores de tipo geométrico ordenados (OWA) con valores como:

$$\hat{F}_i(t) = \frac{m_{(j1)+c} m_{(j1)+c} \dots + c^{k-1} m_{(jk)}}{1+c+\dots+c^{k-1}} \quad (2.32)$$

Para esta última expresión, la constante " c ", se elige mediante iteraciones y los valores $m_{(jk)}$ están ordenados de manera ascendente.

Como se vio existen diversas variantes y combinaciones, para el uso de los promedios generalizados, pero los autores no contienden en la comparación o mejor dicho la elección de cual operador es el óptimo.

2.4. Cadenas de Markov

Las cadenas de Markov fueron introducidas por el matemático ruso Andrey Markov alrededor de 1905. Su intención era crear un modelo probabilístico para analizar la frecuencia con la que aparecen las vocales en poemas y textos literarios.

El éxito del modelo propuesto por Markov radica en que es lo suficientemente complejo como para describir ciertas características no triviales de algunos sistemas, pero al mismo tiempo es lo suficientemente sencillo para ser analizado matemáticamente. Las cadenas de Markov pueden aplicarse a una amplia gama de fenómenos científicos y sociales, y se cuenta con una teoría matemática extensa al respecto.



Ilustración 14. Andrey Markov
(Rusia, 1856-1922)

Definición 2.19. (Proceso Markoviano) Sea el proceso estocástico $\{X(t); t \in T\}$, se establece que es un proceso estocástico de Markov si $\forall n \in \mathbb{N}$ y todo conjunto de índices $t_1 < t_2 < \dots < t_n$, se cumple lo siguiente:

$$P[X(t_n) \leq x_n | X(t_1) = x_1; \dots; X(t_{n-1}) = x_{n-1}] = P[X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}]$$

Se pueden puntualizar en varios conceptos y propiedades sobre cadenas de Markov, pero solo se detallaran los que se enmarcan para el tipo de análisis a realizarse. Referencias más amplias sobre cadenas de Markov se exponen en trabajos como los de Grimmett & Stirzaker (Grimmett & Stirzaker, 1993), Cox & Miller (1965), Marquiegui (Marquiegui, 1994), entre otros más por citar.

Al tratarse de una familia de variables aleatorias, es necesario centrarnos y enmarcarnos dentro de un tipo particular en los procesos de Markov, denominados cadenas de Markov en parámetro discreto, o simplemente cadena de Markov, una clasificación general en base a su parámetro "t", y el espacio de estados $\{E(k)\}$ se detalla en el siguiente cuadro.

TABLA 2. CLASIFICACION DE PROCESOS DE MARKOV

		Espacio de los estados:	
		Discreto	Continuo
Parámetro	Discreto	Cadena de Markov (par. discreto)	Proceso de Markov
	Continuo	Cadena de Markov (par. continuo)	Proceso de difusión

Centrándonos exclusivamente en una cadena de Markov, la probabilidad de encontrarse en el estado j-esimo, para un tiempo²⁴ determinado "t" esta denotada por: $\theta_j^{(t)} = P[X_{(t)} = j]$, y en forma más general para los k-estados se define como un vector:

²⁴ De ahora en adelante el parámetro que los procesos estocásticos emplean, se ha de considerar como el tiempo, pero este hecho se particulariza solo para el desarrollo del tema.

$$\theta^{(t)} = [\theta_1^{(t)}; \theta_2^{(t)}; \dots; \theta_k^{(t)}] \quad t = 0, 1, 2, \dots$$

Cumpliendo además las propiedades de eventos exhaustivos, esto quiere decir que: $\theta_j^{(t)} \geq 0$ y $\sum_j \theta_j^{(t)} = 1$.

El centro de aplicación en una cadena de Markov, radica en el análisis y manejo de las *probabilidades de transición* de un estado a otro para determinados tiempos. Estas probabilidades se denotan por:

$$\theta_{i,j}^{(s,t)} = P[X_{(t)} = j | X_{(s)} = i] \quad \text{con } t > s$$

Sin embargo existe cierto grado de complejidad en el manejo de esta última expresión ya que depende de dos periodos de referencia, pero el concepto *cadena de Markov homogénea* facilita el cálculo y manejo de esta última expresión, reduciéndola a:

$$\theta_{i,j}^{(0,\tau)} = P[X_{(\tau)} = j | X_{(0)} = i] \quad \text{con } \tau > 0$$

Para esta última ecuación solo importa el número de etapas que suceden desde un estado \mathbb{E}_j en el tiempo "0", al instante " τ " en donde se encontrara en el estado \mathbb{E}_i , variando a: $\tau = 1, 2, 3, \dots$ se obtendrán las probabilidades de transición en 1, 2 y 3 etapas respectivamente.

2.4.1. Matriz de transición

Con el manejo de notación matricial, se define la matriz de transición para "t" etapas mediante lo siguiente:

$$\Theta^{(t)} = \begin{bmatrix} \theta_{1,1}^{(t)} & \theta_{1,2}^{(t)} & \dots & \theta_{1,k}^{(t)} \\ \theta_{2,1}^{(t)} & \theta_{2,2}^{(t)} & \dots & \theta_{2,k}^{(t)} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{k,1}^{(t)} & \theta_{k,2}^{(t)} & \dots & \theta_{k,k}^{(t)} \end{bmatrix} \quad (2.33)$$

El caso particular para $t = 1$ origina la matriz de transición en 1 etapa y está es la de mayor interés, debido a que el proceso de estimación se centrara sobre los parámetros que la misma encierra.

$$\Theta = \begin{bmatrix} \theta_{1,1} & \theta_{1,2} & \dots & \theta_{1,k} \\ \theta_{2,1} & \theta_{2,2} & \dots & \theta_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{k,1} & \theta_{k,2} & \dots & \theta_{k,k} \end{bmatrix} \quad (2.34)$$

Recalcando que los elementos de esta matriz deben de cumplir: $\theta_{i,j} \geq 0$ y $\sum_j \theta_{i,j} = 1$; denominándola al mismo tiempo como una *Matriz estocástica*.

En inferencia estadística existen varios métodos para proceso de estimación de parámetros, algunos enmarcados en la línea clásica de la estadística como son: máxima verosimilitud, método de momentos, mínimos cuadrados y otros en líneas de la estadística bayesiana, simulación, etc.

Sin embargo nos centraremos concretamente en dos tipos de métodos: máxima verosimilitud y el método bayesiano, poniendo además énfasis en aspectos que los estimadores generados por estos métodos deberán de cumplir, como lo son la insesgadez, suficiencia, información y los más pertinentes.

2.4.2. Método de Estimación por Máxima Verosimilitud

Sea una muestra de tamaño "n" la cual nos permite elaborar la siguiente matriz de datos observados $\{n_{i,j}\}$, dicha matriz se denota por "H", y cada $n_{i,j}$ representa el numero de observaciones que pasan del estado "i" al estado "j".

$$H = \begin{bmatrix} n_{1,1} & n_{1,2} & \cdots & n_{1,k} \\ n_{2,1} & n_{2,2} & \cdots & n_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ n_{k,1} & n_{k,2} & \cdots & n_{k,k} \end{bmatrix} \quad (2.35)$$

Para este conjunto de valores observados que son exhaustivos es decir: $\sum_{i,j} n_{i,j} = n$. Ahora bien se construyendo la función de máxima verosimilitud, y posteriormente al tomar logaritmo se obtiene:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^k \prod_{j=1}^k \theta_{i,j}^{n_{i,j}} \\ &= \ln[\prod_{i=1}^k \prod_{j=1}^k \theta_{i,j}^{n_{i,j}}] \\ &= \sum_{i=1}^k \sum_{j=1}^k n_{i,j} * \ln(\theta_{i,j}) \end{aligned} \quad (2.36)$$

Esta última ecuación nos lleva a una incoherencia ya que al tomar derivadas llegamos a:

$$\begin{aligned} \frac{\partial \log(L)}{\partial \theta_{i,j}} &= \frac{n_{i,j}}{\theta_{i,j}} \\ \frac{\partial \log(L)}{\partial \theta_{i,j}} &= \frac{n_{i,j}}{\theta_{i,j}} = 0 \quad \Rightarrow \hat{\theta}_{i,j} = \infty \end{aligned}$$

Dicho resultado no tiene sentido, pero con la ayuda de la incorporación de los multiplicadores de Lagrange, en un conjunto de k ecuaciones con suma constante e igual a 1, equivalente a: $\sum_j \theta_{i,j} = 1$. Lleva a plantear la nueva función log-verosímil como:

$$\ln(L(\theta)) = \sum_{i=1}^k \sum_{j=1}^k n_{i,j} * \ln(\theta_{i,j}) + \sum_{i=1}^k \lambda_i * (1 - \sum_j \theta_{i,j})$$

Al tomar derivadas parciales sobre $\theta_{i,j}$ y considerando ahora a los multiplicadores λ_i , se obtiene:

$$\frac{\partial \log(L)}{\partial \theta_{i,j}} = \frac{n_{i,j}}{\theta_{i,j}} - \lambda_i \quad \Rightarrow \quad \theta_{i,j} = \frac{n_{i,j}}{\lambda_i}$$

Pero para el i -ésimo multiplicador de Lagrange λ_i genera la ecuación $\sum_j \theta_{i,j} = 1$, y al sumamos los términos obtenidos respecto a " j " y sumarlos obteniéndose:

$$\begin{aligned} \sum_j \theta_{i,j} &= \sum_j \frac{n_{i,j}}{\lambda_i} \\ &= \frac{1}{\lambda_i} \sum_j n_{i,j} \\ 1 &= \frac{1}{\lambda_i} \sum_j n_{i,j} \quad \Rightarrow \quad \lambda_i = \sum_j n_{i,j} \end{aligned}$$

Esta última expresión lo reemplazamos en la penúltima ecuación, hallando finalmente se obtiene la matriz compuesta por:

$$\begin{aligned} \hat{\theta}_{i,j} &= \frac{n_{i,j}}{\sum_j n_{i,j}} \\ \hat{\theta} &= \{\hat{\theta}_{i,j}\} \end{aligned} \quad (2.37)$$

El anterior método además sirve para estimar estructuras de Markov de orden mayor, concretamente de la forma: $\Theta^{(t)}$; $t > 1$, denominadas *observaciones a intervalos de tiempo iguales*, en las cuales el tiempo de transición es mayor a la unidad. Posteriormente al cálculo del estimador máximo verosímil $\hat{\Theta}^{(t)}$ se realiza la descomposición:

$$\hat{\Theta}^{(t)} = B \Lambda^{(t)} B^{-1}$$

Bajo el supuesto que la matriz $\hat{\Theta}^{(t)}$ es diagonalizable en términos de " B " y " $\Lambda^{(t)}$ ", de la ecuación anterior, se puede hallar de manera directa la matriz $\hat{\Theta}$, la cual será la raíz de orden " t ", es decir:

$$\hat{\Theta} = B [\Lambda^{(t)}]^{1/t} B^{-1}$$

Método de máxima verosimilitud: Propiedades

Fácilmente se verifica que para un tamaño de muestra grande, la distribución marginal de $\hat{\theta}_{i,j}$ sigue una ley asintóticamente normal: $\hat{\theta}_{i,j} \sim \mathcal{N}(\theta_{i,j}; \frac{\theta_{i,j}(1-\theta_{i,j})}{\sum_j n_{i,j}})$ y por lo tanto se puede verificar sin mucha complejidad, propiedades de este estimador como son el sesgo, consistencia, eficiencia y otros aspectos más.

Sin embargo la distribución conjunta de $\{\hat{\theta}_{i,1}, \hat{\theta}_{i,2}, \dots, \hat{\theta}_{i,k}\}$ no es precisamente una distribución normal multivariante, ya que para estas variables se tiene la restricción de

suma constante e igual a la unidad: $\sum_j \hat{\theta}_{i,j} = 1$, y al ser estas el resultado de una transformación de una ley Multinomial de $\{n_{i,1}, n_{i,2}, \dots, n_{i,k}\}$, se deduce que:

$$V[\hat{\theta}_i] = \begin{bmatrix} \omega_{1,1} & \omega_{1,2} & \cdots & \omega_{1,k} \\ \omega_{2,1} & \omega_{2,2} & \cdots & \omega_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{k,1} & \omega_{k,2} & \cdots & \omega_{k,k} \end{bmatrix}$$

Donde cada término $\omega_{l,m}$ está definido por:

$$\omega_{l,m} = \begin{cases} \frac{\theta_{i,l}(1-\theta_{i,l})}{\sum_j n_{i,j}} & l = m \\ -\frac{\theta_{i,l}\theta_{i,m}}{\sum_j n_{i,j}} & l \neq m \end{cases}$$

Método de máxima verosimilitud: Generalización

Todos los conceptos mencionados anteriormente son parte del modelo expuesto por Anderson (Anderson & Goodman, 1955). Donde plantean una función máximo verosímil mediante las siguientes definiciones:

- Sea el tiempo de observación dado por $t = 0, 1, \dots, T$.
- Sea $p_{ij}(t)$ con $i, j = 1, 2, \dots, k$ y $t = 1, 2, \dots, T$; la probabilidad de estar en el estado "j" para el instante "t", dado que estuvo en el estado "i" para el instante "t - 1".
- Contemplando además el hecho de que $p_{ij}(t)$ es una probabilidad no necesariamente estacionaria.
- Dado un conjunto de valores observados $n_{ij}(t)$ con $i, j = 1, 2, \dots, k$ y $t = 1, 2, \dots, T$, contemplando ahora un agregado de $k^2 T$ valores numéricos.

Estos aspectos plasmados en la función de verosimilitud dada por:

$$L(\theta) = \prod_{t=1}^T [\prod_{i,j} p_{ij}(t)^{n_{ij}(t)}] \quad (2.38)$$

Y bajo el mismo procedimiento anterior al del régimen estacionario que se mostro en un inicio, los estimadores máximos verosímiles están dados por:

$$\hat{p}_{i,j}(t) = n_{i,j}(t) / \sum_j n_{i,j}(t) \quad (2.39)$$

Como se ve esta generalización del método de máxima verosimilitud, viene propuesto en el caso de que la cadena de Markov no sea homogénea, y posteriormente se detalla la elaboración de un test que verifique dicha hipótesis.

Prueba de Homogeneidad para Cadenas de Markov

El test de *homogeneidad* que se maneja, está en base a la extensión del teorema de Cramer o el de Neyman como ellos mismos lo citan, basándose que la siguiente expresión:

$$\lambda = -2 \ln \left(\prod_t \prod_{i,j} \left[\frac{\hat{p}_{ij}}{\hat{p}_{ij}(t)} \right]^{n_{ij}(t)} \right) \quad (2.42)$$

Denominado también *test log-ratio*, esta cantidad tiene una distribución Chi-cuadrado: χ_v^2 con $v = k(k-1)(T-1)$ grados de libertad, y rechazando la hipótesis nula de homogeneidad de la cadena de Markov para cierto nivel de significancia.

Otra alternativa también fue plateada, usando las distribuciones multinomiales independientes, con el uso de cada muestra de estado inicial i -esimo:

$$\chi_\eta^2 = \sum_{t,j} \frac{[\sum_j n_{i,j}(t)] * [\hat{p}_{ij}(t) - \hat{p}_{ij}]^2}{\hat{p}_{ij}} \quad (2.43)$$

Siguiendo esta una distribución Chi-cuadrado: χ_η^2 con $\eta = (k-1)(T-1)$ grados de libertad

2.4.3. Método de Estimación Bayesiano

En la estadística clásica también denominada estadística frecuentista, se asume que nuestros datos x_1, x_2, \dots, x_n siguen una distribución $f(x|\theta)$ para un parámetro fijo pero desconocido θ , realizando de esta manera inferencia sobre el valor de dicho parámetro.

En la estadística bayesiana se asume que el parámetro no es exactamente una variable aleatoria, pero se le asigna una función de distribución, estableciendo así una distribución de este condicionado a datos observados, y antes de realizar un experimento se ha de establecer una función $\pi(\theta)$ denominada densidad priori.

Una vez observado los datos, se actualizara nuestra información sobre θ en la nueva función denominada posteriori, con el uso de la regla de Bayes en donde se describirá mejor el comportamiento del parámetro:

$$\pi(\theta|x) = \frac{f(x|\theta) \pi(\theta)}{\int f(x|\theta) \pi(\theta) d\theta}$$

En donde $\pi(\theta)$ es la densidad priori, $f(x|\theta)$ denota la función verosímil $L(\theta)$ y como el denominador no depende del parámetro θ se puede manejar más fácilmente con el concepto de proporcionalidad mediante:

$$\pi(\theta|x) \propto f(x|\theta) \times \pi(\theta)$$

Antes de situarnos concretamente en el campo de la inferencia bayesiana para la estimación de los parámetros de una cadena de Markov, es necesario recalcar las siguientes definiciones en el área de la estadística bayesiana.

Definición 2.20. (Distribuciones conjugadas) Sea \mathcal{F} una clase de distribuciones muestrales denotadas como $f(x|\theta)$, y sea \mathcal{P} una clase de distribuciones a priori expresadas como $\pi(\theta)$ para el parámetro θ , se afirma que \mathcal{P} es conjugada con \mathcal{F} si:

$$\pi(\theta|x) \in \mathcal{P} \quad \forall \quad f(x|\theta) \in \mathcal{F} \wedge \pi(\theta) \in \mathcal{P}$$

Este es si bien un método muy conocido que brinda ventajas considerables en su manejo, pero aun depende de otros parámetros para poder confeccionar la distribución priori, que obviamente deberían ser asumidos por peritos convenientes, dicho esto se presenta una alternativa a continuación y que fácilmente se extiende al caso multivariante (Box & Tiao, 1992).

Definición 2.21. (Distribución a Priori de Jeffreys) La distribución a priori de Jeffreys está dada por: $\pi(\theta) = \sqrt{I(\theta)}$ donde $I(\theta)$ es la información de Fisher: $I(\theta) = -E \left[\frac{d^2}{d\theta^2} \log f(x|\theta) \right]$.

Existirán diversos estimadores dentro de la gama de técnicas bayesianas por aplicar, dado que se ha de elegir distintos tipos de funciones priori para los parámetros de interés. Es así que a continuación se expone una introducción para datos binarios y posteriormente la distribución multivariante más adecuada dentro del campo de las *distribuciones conjugadas*.

Distribución Beta

Antes de empezar a exponer el caso multivariante, abordaremos de manera esencial el proceso de estimación bayesiana del caso univariado, el uso de la distribución a priori conjugada, es la distribución beta denotada por "Be" y la cual Lindley (Lindley, 1964) desarrolla.

Definición 2.22. (Distribución Beta) Sea una variable aleatoria $X \in \mathbb{R}$ y sean los parámetros $\alpha, \beta > 0$; la variable aleatoria X tiene una distribución beta denotado como $X \sim Be(\alpha; \beta)$ si su función de densidad viene dada por:

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad 0 < x < 1 \quad (2.40)$$

La distribución beta es suficientemente flexible como para recoger la mayor parte de las posibles creencias del experto. Básicamente cualquier distribución unimodal acotada entre los valores $[0; 1]$ será muy aproximada por una distribución beta. Como vemos en

la figura a continuación, las densidades beta adoptan formas absolutamente distintas para diferentes valores de los parámetros.

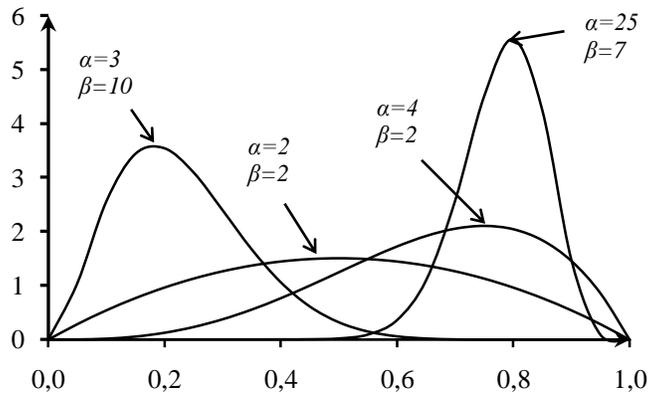


Ilustración 15. Densidades beta para distintos valores de α y β

Teorema 2.1. (Distribución Beta) Sea $\theta \in (0; 1)$ un parámetro de interés sobre el que deseamos hacer inferencia. Suponiendo que para θ asignamos una densidad a priori $Be(\alpha; \beta)$ y dada una muestra de tamaño “n” se han observado “x” éxitos de tipo Bernoulli (θ). Entonces, la densidad posteriori de θ , es también una densidad beta con $Be(\alpha + x; \beta + n - x)$

Demostración: la función de densidad priori está dada por $\pi(\theta) \propto \theta^{\alpha-1}(1 - \theta)^{\beta-1}$ al tomar la función verosímil $l(x|\theta) = \theta^x(1 - \theta)^{n-x}$, y reemplazando obtenemos:

$$\begin{aligned} \pi(\theta|x) &\propto l(x|\theta) \times \pi(\theta) \\ &\propto \theta^x(1 - \theta)^{n-x} \theta^{\alpha-1}(1 - \theta)^{\beta-1} \\ &\propto \theta^{x+\alpha-1}(1 - \theta)^{n-x+\beta-1} \\ \therefore \pi(\theta|x) &\sim Be(\alpha + x; \beta + n - x) \end{aligned}$$

Al considerar esta distribución posteriori beta se analizan algunos indicadores como su valor esperado $E[\theta|x]$ y su varianza $V[\theta|x]$, los cuales se han de descomponer parcialmente.

$$\begin{aligned} E[\theta|x] &= \hat{\alpha} / (\hat{\alpha} + \hat{\beta}) \\ &= (\alpha + x) / (\alpha + \beta + n) \\ &= \alpha / (\alpha + \beta + n) + x / (\alpha + \beta + n) \\ &= \alpha(\alpha + \beta) / [(\alpha + \beta)(\alpha + \beta + n)] + x n / n(\alpha + \beta + n) \end{aligned}$$

Si denotamos $w_n = n / (\alpha + \beta + n)$ como un ponderador de información posteriori.

$$\begin{aligned} &= [(\alpha + \beta) / (\alpha + \beta + n)] \alpha / (\alpha + \beta) + w_n * x / n \\ &= [(\alpha + \beta + n - n) / (\alpha + \beta + n)] E[\theta] + w_n * x / n \\ &= [1 - n / (\alpha + \beta + n)] E[\theta] + w_n * x / n \\ &= (1 - w_n) E[\theta] + w_n * x / n \end{aligned}$$

No es complejo el hecho de verificar que este ponderador w_n asintóticamente tiende a 1, debido a esto la esperanza posteriori también cumple lo siguiente:

$$\begin{aligned}\lim_{n \rightarrow \infty} E[\theta|x] &= \lim_{n \rightarrow \infty} [(1 - w_n)E[\theta] + w_n * x/n] \\ &= \lim_{n \rightarrow \infty} (1 - w_n)E[\theta] + \lim_{n \rightarrow \infty} w_n * x/n \\ &= (1 - 1)E[\theta] + \lim_{n \rightarrow \infty} w_n * \lim_{n \rightarrow \infty} x/n \\ &= 0 + 1 * \lim_{n \rightarrow \infty} x/n \\ \lim_{n \rightarrow \infty} E[\theta|x] &= \theta\end{aligned}$$

De manera similar se descompone el término de la varianza $V[\theta|x]$, llegando a tender al valor de: $\theta(1 - \theta)/n$ al igual que el estimador que se obtiene por el método de máxima verosimilitud.

Elección de la Distribución Priori

Si bien existirán opiniones distintas sobre la elección de la distribución priori más adecuada, e incluso ya habiendo asumido una distribución beta en sí, llega la incógnita de que valores de " α " y " β " que influirán considerablemente en la obtención de la distribución posteriori.

Con el uso de la definición de *Jeffreys* (Jeffreys, 1961), se llega a una distribución beta de parámetros: $\mathcal{B}e(\alpha = 1/2; \beta = 1/2)$, esta aseveración priori para el caso de estimar el parámetro θ de experimentos Bernoulli, se construye de la siguiente manera:

Demostración: la función de probabilidad está dada por $p(x_i|\theta) = \theta^{x_i}(1 - \theta)^{1-x_i}$ obtener la información de Fisher definida como:

$$I[\theta] = -E \left[\frac{d^2 \ln(p(x_i|\theta))}{d\theta^2} \right]$$

Tomando logaritmo a la función de probabilidad se ha de obtener: $\ln[p(x_i|\theta)] = x_i \ln(\theta) + (1 - x_i)\ln(1 - \theta)$

$$\begin{aligned}\ln(p(x_i|\theta)) &= x_i \ln(\theta) + (1 - x_i)\ln(1 - \theta) \\ \frac{d \ln(p(x_i|\theta))}{d\theta} &= \frac{x_i}{\theta} - \frac{1-x_i}{1-\theta} \\ \frac{d^2 \ln(p(x_i|\theta))}{d\theta^2} &= -\frac{x_i}{\theta^2} - \frac{1-x_i}{(1-\theta)^2}\end{aligned}$$

Como son distribuciones Bernoulli se cumple que $E[x_i] = \theta$ en esta última expresión, y a tomar valores esperados en la anterior ecuación:

$$\begin{aligned}E \left[\frac{d^2 \ln(p(x_i|\theta))}{d\theta^2} \right] &= E \left[-\frac{x_i}{\theta^2} - \frac{1-x_i}{(1-\theta)^2} \right] \\ &= -\frac{1}{\theta^2} E[x_i] - \frac{1}{(1-\theta)^2} E[1 - x_i]\end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{\theta^2}\theta - \frac{1}{(1-\theta)^2}(1-\theta) \\
&= -\frac{1}{\theta(1-\theta)}
\end{aligned}$$

Finalmente la información de Fisher es $I[\theta] = 1/\theta(1-\theta)$, *Harold Jeffreys* establece que una distribución poco informativa para θ se define en función a la información de *Fisher*, bajo la siguiente relación:

$$\pi(\theta) \propto \sqrt{I[\theta]}$$

Con lo cual la información de Fisher es: $\sqrt{I[\theta]} = \sqrt{1/\theta(1-\theta)} = \theta^{-1/2}(1-\theta)^{-1/2}$, correspondiendo de esta manera a una distribución beta $\mathcal{Be}(\alpha = 1/2; \beta = 1/2)$.

Elección del Estimador Puntual

Se determino la distribución posteriori de θ y analizo el valor esperado de esta, pero con $\pi(\theta|x)$ se puede calcular medidas centrales, de posición, de variación, etc., sin embargo esto no es muy imprescindible en el contexto de desarrollo del tema.

La elección del mejor estimador de θ , se ha de justificar mediante el uso de una función de pérdida denotada por $L(\theta; \tilde{\theta})$, donde $\tilde{\theta}$ es el estimador de θ , y el cual se justificara al minimizar la pérdida cuantificada en $R_{\tilde{\theta}}(\theta)$, denominado *pérdida esperada posterior*, en resumen:

$$R_{\tilde{\theta}}(\theta) = \int_{-\infty}^{+\infty} L(\theta; \tilde{\theta}) \pi(\theta|x) d\theta$$

Centrándonos en la función de pérdida más fundamental, la cual recae en el uso la de pérdida cuadrática, $L(\theta; \tilde{\theta}) = (\theta - \tilde{\theta})^2$, con lo cual la la siguiente expresión:

$$R_{\tilde{\theta}}(\theta) = \int_{-\infty}^{+\infty} (\theta - \tilde{\theta})^2 \pi(\theta|x) d\theta$$

Para estas no es complejo que el valor mínimo se alcanza si y solo si $\tilde{\theta} = E[\theta|x]$, es decir se usa la esperanza posteriori como estimador de θ , otras opciones son la moda o la mediana de la distribución posteriori.

Distribución Dirichlet

Las denominadas distribuciones conjugadas se consideran una herramienta muy poderosa, para la comprensión de la información, sobre todo en la comparación del modelo. La forma exacta de la distribución de Dirichlet, equilibra de manera adecuada la información de los datos recolectados para la obtención de una función posteriori.

Definición 2.20. (Distribución Dirichlet) Sea un vector aleatorio $\vec{X} = (X_1, X_2, \dots, X_k) \in \mathbb{R}_+^k$ donde se verifica que $\sum_i X_i = 1$, y sea el vector de parámetros

$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ $\alpha_i > 0$; el vector aleatorio \vec{X} tiene una distribución Dirichlet denotado como $\vec{X} \sim Dir(\alpha)$ y si su función de densidad viene dada por:

$$f(\vec{x}; \alpha) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \prod_i x_i^{\alpha_i - 1} \quad \forall x_i \in (0; 1) \quad (2.40)$$

La representación grafica de esta densidad multivariante es limitada, pero a continuación se particulariza para el caso en 3 dimensiones $k = 3$, mediante el uso de curvas de nivel ya que da una mejor perspectiva, y cabe recordar que la restricción de suma unitaria permite excluir la 3ra componente que está representada por el origen.

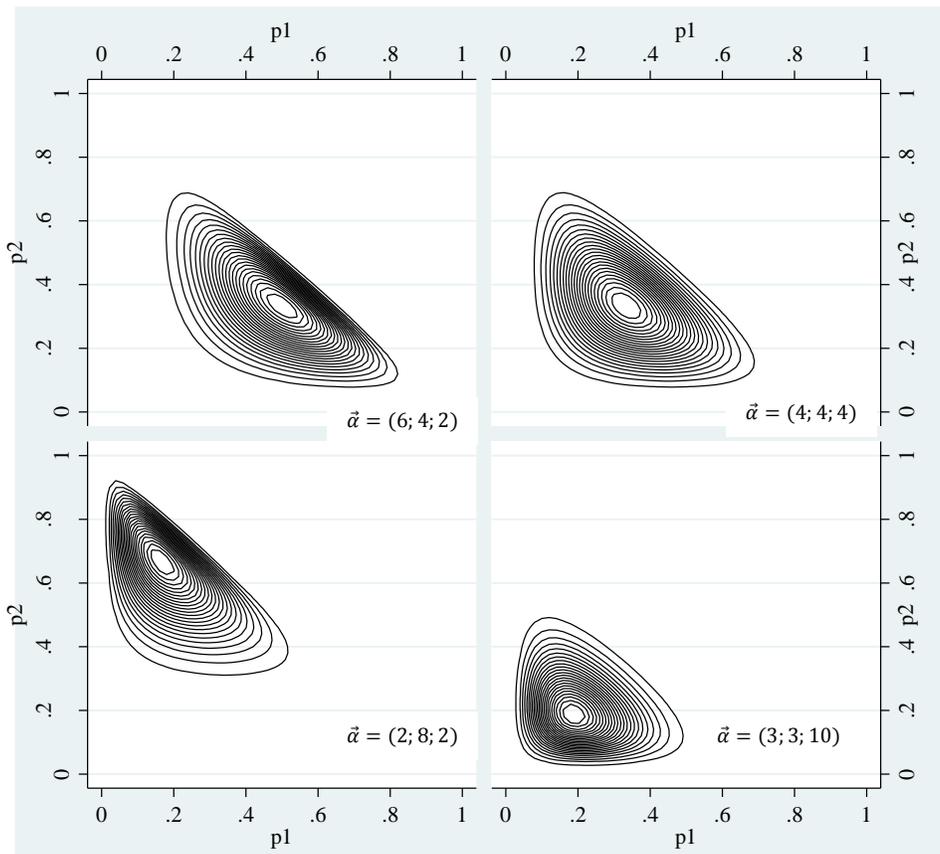


Ilustración 16. Curvas de nivel para distintas densidades Dirichlet

Algunos autores denominan $\alpha_0 = \sum_i \alpha_i$, para el manejo de las distribuciones marginales de cada X_i , las cuales se comportan mediante una distribución beta: $X_i \sim Be[\alpha_i; \alpha_0 - \alpha_i]$ con valor esperado igual a α_i/α_0 , este hecho no es muy complicado de verificar, y con esta prueba se puede construir la matriz de varianzas y covarianzas definida como:

$$V[\vec{X}] = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,k} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k,1} & \sigma_{k,2} & \cdots & \sigma_{k,k} \end{bmatrix}$$

Donde cada término $\sigma_{i,j}$ está definido por:

$$\sigma_{i,j} = \begin{cases} \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)} & i = j \\ -\frac{\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)} & i \neq j \end{cases}$$

Con estas aclaraciones, consideramos una muestra aleatoria proveniente de una distribución multinomial con vector de parámetros $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ tales que $\sum_j \theta_j = 1$, dicha muestra de tamaño "n" resumida en el vector $H = (n_1, n_2, \dots, n_k)$ con $\sum_j n_j = n$, generando:

$$\begin{aligned} f(\theta; H) &= \pi(\theta) * L(H|\theta) \\ &= \frac{\Gamma(\sum_j \alpha_j)}{\prod_j \Gamma(\alpha_j)} \prod_j \theta_j^{\alpha_j - 1} * \prod_j \theta_j^{n_j} \\ &= \frac{\Gamma(\sum_j \alpha_j)}{\prod_j \Gamma(\alpha_j)} \prod_j \theta_j^{\alpha_j + n_j - 1} \end{aligned}$$

Determinada la función conjunta se obtiene al cálculo de la *distribución conjugada* posteriori de $\theta|H$ obteniendo lo siguiente:

$$\begin{aligned} f(\theta|H) &\propto f(\theta; H) \\ &\propto \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \prod_i \theta_i^{\alpha_i + n_i - 1} \\ &\propto \prod_i \theta_i^{\alpha_i + n_i - 1} \\ \therefore \theta|H &\sim \text{Dir}(\tilde{\alpha}) \end{aligned} \quad (2.41)$$

Y de este modo también sigue una distribución Dirichlet, donde los parámetros vienen dados por: $\tilde{\alpha} = (\alpha_1 + n_1, \alpha_2 + n_2, \dots, \alpha_k + n_k)$ y $\tilde{\alpha}_0 = \sum_i \alpha_i + n$, este hecho nos ayudan a determinar el vector de valores esperados compuesto por: $E[\theta_i|H] = (\alpha_i + n_i)/\tilde{\alpha}_0$, de manera más directa este vector es el estimador puntual de "θ", aclarando que se usa la función cuadrática de perdida $L(\theta; \tilde{\theta}) = (\theta - \tilde{\theta})^2$.

Priori multivariante de Jeffreys

Ahora se pondrá en exposición la generalización de la distribución Beta en la distribución a priori de Jeffreys, para el caso de una distribución Dirichlet en muestras multinomiales.

Demostración: la función de probabilidad para una multinomial está dada por:

$$Pr(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k | \theta) = \frac{n!}{\prod_j n_j!} \prod_j \theta_j^{n_j}$$

Para obtener la matriz de información de Fisher $I[\theta]$ primero tomamos logaritmos:

$$\begin{aligned} \ln[p(X|\theta)] &= \ln(n!) - \ln(\prod_j n_j!) + \ln(\prod_j \theta_j^{n_j}) \\ &= \ln(n!) - \ln(\prod_j n_j!) + \sum_j n_j \ln(\theta_j) \end{aligned}$$

Antes de establecer la información de Fisher $I[\theta]$ de manera adecuada, recordamos el hecho que $\theta_k = 1 - \sum_{j < k} \theta_j$ y $n_k = n - \sum_{j < k} n_j$, con lo cual se puede proceder a obtener las correspondientes derivadas de primer y segundo orden escritas como:

$$\begin{aligned} \frac{d \ln[p(X|\theta)]}{d\theta_i} &= \frac{n_i}{\theta_i} - \frac{n - \sum_{j < k} n_j}{1 - \sum_{j < k} \theta_j} \\ \frac{d^2 \ln[p(X|\theta)]}{d\theta_i^2} &= -\frac{n_i}{\theta_i^2} - \frac{n - \sum_{j < k} n_j}{(1 - \sum_{j < k} \theta_j)^2} \\ &= -\frac{n_i}{\theta_i^2} - \frac{n_k}{\theta_k^2} \end{aligned}$$

En esta etapa se debe aclarar que la derivada de segundo orden $d^2/d\theta_i \theta_l$ para índices $i \neq l$ es tan solo el segundo término de la expresión antes mostrada, y con el conocimiento que $E[n_i] = n \theta_i$ para $i = 1, 2, \dots, k$.

$$\begin{aligned} \frac{d^2 \ln[p(X|\theta)]}{d\theta_i^2} &= -\frac{n_i}{\theta_i^2} - \frac{n - \sum_{j < k} n_j}{(1 - \sum_{j < k} \theta_j)^2} \\ E \left[\frac{d^2 \ln[p(X|\theta)]}{d\theta_i^2} \right] &= E \left[-\frac{n_i}{\theta_i^2} - \frac{n - \sum_{j < k} n_j}{(1 - \sum_{j < k} \theta_j)^2} \right] \\ &= -\frac{1}{\theta_i^2} E[n_i] - \frac{1}{\theta_k^2} E[n_k] \\ &= -\frac{1}{\theta_i^2} n\theta_i - \frac{1}{\theta_k^2} n\theta_k \\ &= -\frac{n}{\theta_i} - \frac{n}{\theta_k} \\ &= -n \left(\frac{1}{\theta_i} + \frac{1}{\theta_k} \right) \\ &= -n \left(\frac{\theta_i + \theta_k}{\theta_i \theta_k} \right) \end{aligned}$$

La matriz $S[\theta]$ se denomina score de θ , y la misma está conformada por los elementos $a_{ii} = -n \left(\frac{\theta_i + \theta_k}{\theta_i \theta_k} \right)$ y $a_{ij} = a_{ji} = -\frac{n}{\theta_k}$, y finalmente se llega a:

$$J[\theta] = \begin{bmatrix} -n \left(\frac{\theta_1 + \theta_k}{\theta_1 \theta_k} \right) & -\frac{n}{\theta_k} & \cdots & -\frac{n}{\theta_k} \\ -\frac{n}{\theta_k} & -n \left(\frac{\theta_2 + \theta_k}{\theta_2 \theta_k} \right) & \cdots & -\frac{n}{\theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{n}{\theta_k} & -\frac{n}{\theta_k} & \cdots & -n \left(\frac{\theta_{k-1} + \theta_k}{\theta_{k-1} \theta_k} \right) \end{bmatrix}$$

La información de Fisher $I[\theta]$ la cual es un escalar, se define como el determinante de esta matriz mediante $I[\theta] = |J[\theta]|$.

$$I[\theta] = \left| \begin{bmatrix} -n \left(\frac{\theta_1 + \theta_k}{\theta_1 \theta_k} \right) & -\frac{n}{\theta_k} & \cdots & -\frac{n}{\theta_k} \\ -\frac{n}{\theta_k} & -n \left(\frac{\theta_2 + \theta_k}{\theta_2 \theta_k} \right) & \cdots & -\frac{n}{\theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{n}{\theta_k} & -\frac{n}{\theta_k} & \cdots & -n \left(\frac{\theta_{k-1} + \theta_k}{\theta_{k-1} \theta_k} \right) \end{bmatrix} \right|$$

Factorizando el término $-n/\theta_k$ de cada fila, se reduce a:

$$I[\theta] = \left| \begin{bmatrix} \frac{\theta_1 + \theta_k}{\theta_1} & 1 & \cdots & 1 \\ 1 & \frac{\theta_2 + \theta_k}{\theta_2} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & \frac{\theta_{k-1} + \theta_k}{\theta_{k-1}} \end{bmatrix} \right| (-n/\theta_k)^{k-1}$$

Bajo un proceso adecuado²⁵ se puede verificar que el determinante de la matriz $Score[\theta]$ es igual a $I[\theta] = n^{k-1} \theta_1^{-1} \theta_2^{-1} \dots \theta_k^{-1}$, algo paradójico es el hecho en el cual se llega al mismo resultado, al omitir la dependencia de θ_k bajo la restricción $\theta_k = 1 - \sum_{j < k} \theta_j$.

$$\begin{aligned} \pi(\theta) &\propto \sqrt{I[\theta]} \\ \pi(\theta) &\propto \sqrt{n^{k-1} \theta_1^{-1} \theta_2^{-1} \dots \theta_k^{-1}} \\ \pi(\theta) &\propto \theta_1^{-1/2} \theta_2^{-1/2} \dots \theta_k^{-1/2} \end{aligned}$$

Y al igual que el caso binario esta distribución encaja en su similar multivariante, es decir en la densidad Dirichlet: $Dir(\alpha_1 = 1/2, \alpha_2 = 1/2, \dots, \alpha_k = 1/2)$, la representación grafica de esta densidad poco informativa se muestra a continuación.

²⁵ Se usa el método de inducción, el mismo se desglosa ampliamente en anexos.

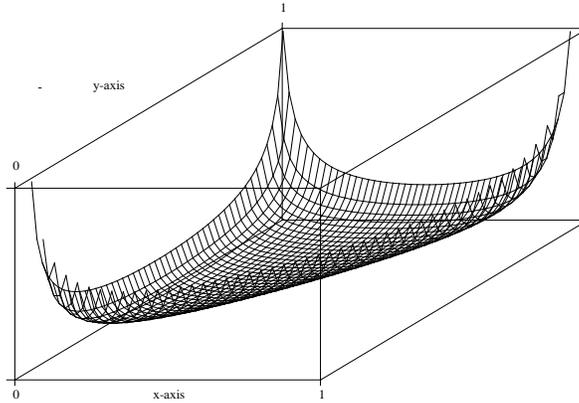


Ilustración 17. Densidad priori poco informativa $Dir(\alpha_1 = \alpha_2 = \alpha_3 = 1/2)$

2.5. Aplicación de Cadenas de Markov

Una grafica sobre una cadena de Markov mediante la representación sagital, lleva a entender casi la misma noción de estructura que Song & Chissom's introdujeron para las series difusas, sin embargo el nexa que une estas dos metodologías debe de ser expuesto más detalladamente.

Sullivan (Sullivan & Woodall, 1994), compararon con la ayuda de una matriz de relaciones de tipo Markoviano el modelo Song & Chissom's, obteniendo de esta manera buenos avances en las series difusas, pero en ese entonces no se realizo la incorporación de la metodología de Chen, ya que esta fue propuesta posteriormente. A continuación se detalla el la secuencia de pasos que ellos usaron:

- Se crean los " v " conjuntos difusos de manera constante, en el espacio de discusión.
- En base a las relaciones de primer orden planteadas en la 2da definición de series de tiempo difusas, se elabora la matriz de datos observados $\{o(i, j)\}$, mediante un método que ellos denominan normalización.
- De la misma manera y con los datos observados $\{o(i, j)\}$ calcula el vector de estados iniciales denominados como $\pi = (\pi_1, \pi_2, \dots, \pi_v)$
- Una vez elaboradas estas matrices expresadas como $\tilde{\theta}$, π y junto a los valores defuzificados de los " v " conjuntos denotados por $\{m_i\}$ se usan como ponderadores los elementos de la i -esima fila para estimar el valor de $\hat{F}_i(t)$.

Nótese que no se manejan conceptos de: relación grupo FLRG, tampoco los agregadores o promedio generalizados, y menos criterios de estimación de la matriz de transición, estas mejoras e incorporaciones al modelo se han de implementan en el capitulo siguiente.

Capítulo III

Marco Práctico

3.1. Preliminares

Tal como se menciono, la metodología de series de tiempo difusas, y más convenientemente el Modelo de Markov, posee tanto ventajas como limitantes. Aclarado este punto se opto por el manejo de una serie temporal, la cual se adecue a la metodología de los modelos ARIMA, brindando así una comparación análoga entre ambas metodologías.

3.1.1. Reforma al Sistema de Reparto

Para finales de 1996, se aprobó en el Congreso de Bolivia la Ley de Reforma del Sistema de Pensiones, con el objetivo de modernizar las disposiciones políticas, económicas y sociales del país (Aponte, Ferrufino, & Jemio, 2008).

Con la intención de cambiar el sistema de pensiones vigente hasta esa fecha, el gobierno había analizado varias alternativas, incluyendo una propuesta para solamente modificar el sistema de reparto vigente, y otras que implicaban un cambio completo del sistema. Finalmente, se optó por una estrategia que reemplazó completamente el antiguo sistema de reparto por uno de capitalización individual.

La Reforma de Pensiones sustituyó el antiguo sistema de reparto simple, por un sistema de capitalización individual, administrado privadamente. Antes de la reforma, el sistema de reparto estaba constituido por un Parte I: La Reforma de Pensiones en Bolivia fondo administrado públicamente (Fondo de Pensiones Básicas, FOPEBA), que cubría el 43 por ciento de la pensión de los jubilados, y un sistema de Fondos Complementarios (FONCOM'S) administrados privadamente por instituciones organizadas y administradas por los mismos trabajadores, en las diferentes ramas de actividad en la que ellos se desempeñaban. Los FONCOM'S cubrían el 57 por ciento adicional de la jubilación percibida por los beneficiarios.

Esta reforma fue ejecutada debido a los problemas que presentaba el antiguo sistema de reparto simple, referidos a:

- i) Transición demográfica, es decir que la población boliviana se encuentra en un proceso de paulatino envejecimiento, por lo que el antiguo sistema iba a colapsar en el largo plazo debido a la insuficiencia de trabajadores activos para financiar a los pasivos;
- ii) Baja relación de contribuyentes sobre beneficiarios, ya que al momento de la reforma existían en promedio 3 trabajadores activos por cada pensionista, cuando el número mínimo requerido por un sistema de reparto para funcionar eficientemente es de 10 a 1.
- iii) Insuficientes años y montos de contribución, ya que el sistema anterior es un sistema de beneficio definido, en el que los beneficiarios obtienen un monto de jubilación determinado, independientemente del número y monto de contribuciones que éstos hubieran realizado durante su vida útil.
- iv) Baja cobertura, debido a que el sistema de reparto solamente cubría a 25.6 por ciento de la población ocupada urbana y prácticamente cero de la población ocupada rural.
- v) Un nivel de reservas insuficiente, ya que éstas ascendían a solamente US\$ 150 millones, equivalentes a 5 por ciento de los pasivos acumulados por el sistema de reparto al momento de la reforma.
- vi) Altos costos de administración, equivalentes en algunos fondos a más del 20 por ciento de las cotizaciones, mientras que la ley establecía como techo máximo un 10 por ciento.

Frente a estos problemas, el objetivo fundamental de la Reforma de Pensiones fue el de mejorar en forma sustancial el funcionamiento del seguro obligatorio de largo plazo en el país, de tal forma que los afiliados pudieran beneficiarse de un sistema que les proporcionara un seguro de vejez adecuado, que les permitiera vivir en forma digna una vez alcanzada la edad de jubilación.

3.1.2. El Nuevo Sistema de Reparto

El nuevo sistema de pensiones implementado en 1997, funciona esencialmente a partir de las contribuciones individuales de cada trabajador a cuentas personales administradas privadamente. Un sistema de capitalización individual, es por definición completamente auto financiable, lo que significa que en el largo plazo el sistema no requerirá transferencias externas provenientes del gobierno. Incluso más importante, el sistema de capitalización individual es un sistema de contribución definido, en contraste con el sistema de reparto que era un sistema de beneficio definido.

Como se describió anteriormente, el sistema de reparto requería que los afiliados realicen un número definido de contribuciones al sistema, después de lo cual estaban facultados para acogerse a retiro y se les garantizaba una renta equivalente al menos al 70 por ciento de su salario por el resto de su vida.

Los beneficios bajo este sistema podían ser por lo tanto independientes del monto acumulado por las contribuciones realizadas por el trabajador. Es importante notar que bajo el sistema de reparto, no existía límite de tiempo para los beneficios financieros que un pensionista podía recibir después de jubilarse, una vez que éste hubiera cumplido con los requisitos que el sistema le exigía, aunque solamente hubiera contribuido durante 15 años.

3.1.3. Las Administradoras de Fondos de Pensiones

Con la aprobación de la Ley 1732 en noviembre de 1996, el nuevo sistema de capitalización individual fue puesto en vigencia, poniendo fin al anterior mediante la transferencia completa de todos los afiliados a los fondos complementarios y el Fondo de Pensiones Básicas se les permitió mantener su estatus legal en forma temporal, con el propósito de evaluar y liquidar los activos de los fondos. Estos activos fueron posteriormente transferidos al Tesoro General de la Nación para cubrir en forma muy parcial, los abultados costos del sistema anterior.

Bajo el nuevo sistema, los fondos de pensiones son administrados por dos *Administradoras de Fondos de Pensiones (AFP'S) privadas*, las cuales cobran un monto fijo mensual de US\$ 2.4 por afiliado; por esta tarifa ellas administran los fondos de retiro de los afiliados, el seguro de invalidez, muerte y riesgo profesional.

Su función más importantes es la administración de los Fondos de Capitalización Individual (FCI), que comprende las contribuciones de los afiliados al sistema de pensiones y el Fondo de Capitalización Colectivo (FCC), que contiene el 50 por ciento del total de las acciones de las empresas capitalizadas, que permanecieron en manos de los bolivianos, después de realizarse la Capitalización de las principales empresas públicas, y fueron transferidas a los fondos de pensiones para financiar el beneficio del Bono Solidario (BONOSOL).

La característica más importante del nuevo sistema es que éste está basado en la idea de un fondo de capitalización individual obligatorio que es completamente autofinanciado y dependiente de las contribuciones y de la rentabilidad del fondo.

El nuevo sistema requiere que los trabajadores afiliados ahorren un 10 por ciento del total de sus ingresos, monto que es depositado en cuentas personales a ser utilizadas en forma directa para financiar el retiro de los individuos. Tanto asalariados como trabajadores por cuenta-propia pueden hacer contribuciones mayores al 10 por ciento, con el objeto de mejorar las condiciones de su retiro o para acceder a un retiro temprano.

3.2. Aplicación del Modelo de Markov Difuso

Previo al manejo de la información, en el apartado anterior se dio una extensa explicación de su procedencia, ahora se encara la descripción, de los datos en sí. El origen serie analizada en su integridad se encuentra disponible en el portal de UDAPE: Unidad de Análisis de Políticas (<http://www.udape.gob.bo/>).

La variable analizada es: “Total de recaudaciones [miles de Bs] por concepto de aportes a las AFP’S”, a periodos mensuales desde Junio de 1997 hasta diciembre del 2011”, esto para el departamento de Potosí, la serie como tal, se ve en el siguiente grafico:

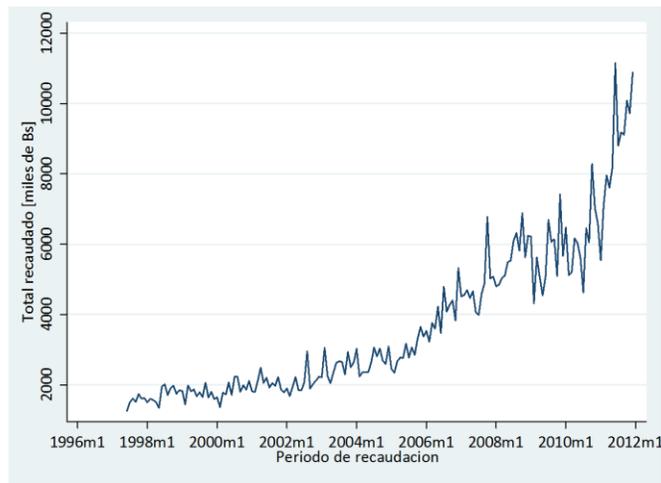


Ilustración 18. Serie de Datos 06-1997 a 12-2011

Claramente se observa una tendencia creciente sobre la serie de datos, sin embargo para fines comparativos con el modelo ARIMA, se usara la transformación $Z_t = \ln(Y_t)$ y con la aplicación de una diferencia, es decir: $W_t = Z_t - Z_{t-1}$, quedando la serie finalmente como:

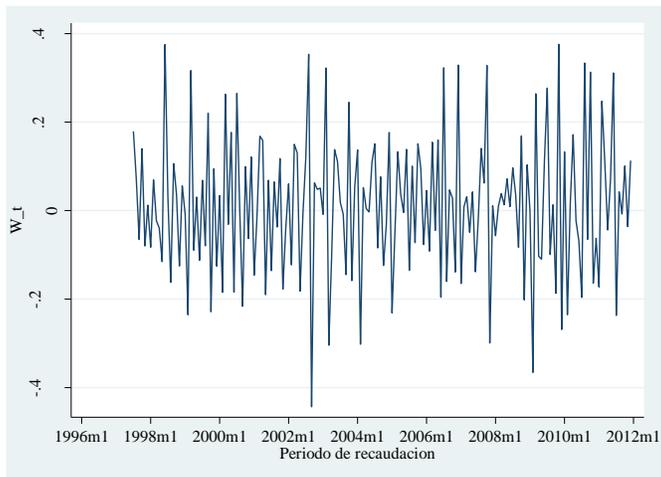


Ilustración 19. Serie de Datos Transformados W_t

3.2.1. Construcción de los Clústers

Para la elaboración de los clusters o conjuntos difusos, se maneja el criterio de los cuantiles (Arutchelvan, Sivasta, & Jaganathan, 2010), donde como se conoce se divide al conjunto de datos en k grupos de igual cantidad de miembros. La grafica que a continuación se muestra, da un panorama sobre dicha clasificación.

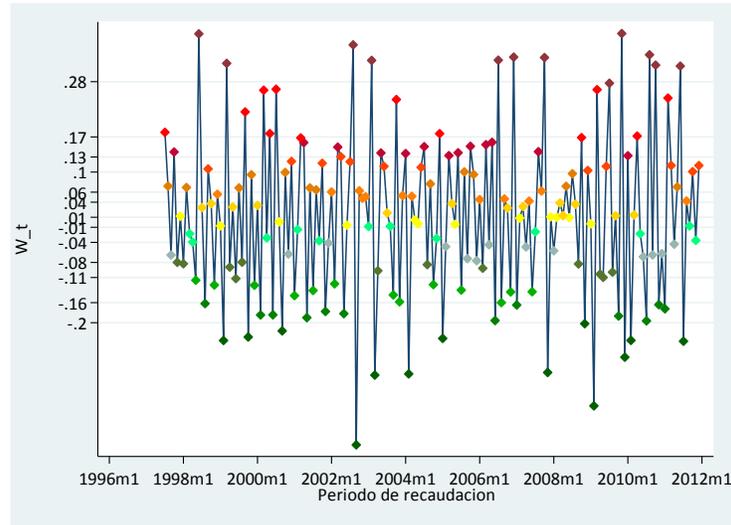


Ilustración 20. Clusters construidos ($k=14$)

Al ser clusters o grupos exhaustivos, poseen cada uno de ellos y de manera aislada sus propios estadísticos. Para este grupo de $k = 14$ clusters, se tiene lo siguiente:

TABLA 3. DESCRIPCIÓN DE LOS CLUSTERS CONFORMADOS

Clusters	Estadísticos							
	n	min	max	p25	p50	p75	mean	sd
A_1	13	-0,44	-0,20	-0,30	-0,24	-0,23	-0,27	0,068
A_2	12	-0,20	-0,16	-0,19	-0,18	-0,17	-0,18	0,012
A_3	13	-0,16	-0,12	-0,14	-0,14	-0,12	-0,14	0,014
A_4	12	-0,11	-0,08	-0,10	-0,09	-0,08	-0,09	0,012
A_5	13	-0,08	-0,04	-0,07	-0,06	-0,05	-0,06	0,012
A_6	12	-0,04	-0,01	-0,03	-0,02	-0,01	-0,02	0,012
A_7	12	-0,01	0,01	0,00	0,00	0,01	0,00	0,007
A_8	13	0,01	0,04	0,02	0,03	0,04	0,03	0,010
A_9	12	0,04	0,06	0,05	0,05	0,06	0,05	0,007
A_{10}	13	0,06	0,10	0,07	0,07	0,10	0,08	0,014
A_{11}	12	0,10	0,13	0,11	0,11	0,12	0,11	0,008
A_{12}	13	0,13	0,16	0,14	0,14	0,15	0,14	0,009
A_{13}	12	0,17	0,26	0,17	0,20	0,26	0,21	0,042
A_{14}	12	0,28	0,38	0,31	0,33	0,34	0,33	0,028

Por ahora la información de estos $k = 14$ clusters, no es relevante, sin embargo es más relevante establecer las relaciones difusas (Chen, 1996) en base al número de elementos que estos contienen, estas mismas se obtienen al obtener un simple cruce de variables, pero entre las variables de clasificación de clusters A_i y A_j . Donde A_j viene del periodo $t - 1$ el cual antecede a A_i del periodo t , esto conlleva a la siguiente tabla:

TABLA 4. RELACIONES DIFUSAS AGRUPADAS

$A_j \rightarrow A_i$	Total	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}	A_{14}
	173	13	12	13	12	13	12	12	13	12	13	12	13	11	12
A_1	13				1	1		1	1	3	2	1	1	1	1
A_2	12					2		2				1	1		3
A_3	13						3		1	3	2			1	1
A_4	12				1			1	2			3	1	2	2
A_5	13		2							3	1	1	4		1
A_6	12			2		1	1						3	1	3
A_7	12	3			1	1		1	2			1	2	1	
A_8	13		3	2	2	1	1	1	1			1			1
A_9	12			2	1		2	2	1	2				1	
A_{10}	13			3	1	4	2	1	1						1
A_{11}	11		2	1		1	1	1	2					1	2
A_{12}	13	2	2	1	2	1			1	1	1	2			
A_{13}	12	3	1	1	1		2	1				1	1	1	
A_{14}	12	5	2	1	2	1			1						

Para poder usar la incorporación de uso de cadenas de Markov nos concentraremos en el manejo de la matriz de datos, que ahora denotaremos como " H ", la cual en resumidas cuentas es:

$$H_{14 \times 14} = \begin{bmatrix} 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 3 \\ \vdots & \vdots & \ddots & \vdots \\ 5 & 2 & \dots & 0 \end{bmatrix}$$

3.2.2. Estimación de la Matriz de Markov

Obtenida la matriz de datos observados $H_{14 \times 14}$, se procede a la estimación para generar la matriz de transición de Markov $\hat{\theta}$, previa a esta fase se hará la incorporación de la distribución priori de θ denotada por θ_0 , los parámetros fueron elegidos según criterios adecuados (Jilani, Burney, & Ardil, 2008), la matriz está dada por:

$$\theta_0 = \begin{bmatrix} 1 & 0,5 & 0 & \dots & 0 & 0 & 0 \\ 0,5 & 1 & 0,5 & \dots & 0 & 0 & 0 \\ 0 & 0,5 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0,5 & 0 \\ 0 & 0 & 0 & \dots & 0,5 & 1 & 0,5 \\ 0 & 0 & 0 & \dots & 0 & 0,5 & 1 \end{bmatrix}$$

El continuar con la estimación en curso, nos sujetamos a lo que se expuso en el capítulo II, donde se demostró que los parámetros de la distribución posteriori para una Dirichlet resultaban como $\tilde{\alpha} = (\alpha_1 + n_1, \alpha_2 + n_2, \dots, \alpha_k + n_k)$. En este caso multiparamétrico que se presenta, no es más que la extensión de los mismos, ya que ahora de manera matricial se maneja $Y = H + \theta_0$.

Constituida esta distribución posteriori, se procede a la estimación bajo una función de pérdida cuadrática $L(\theta; \tilde{\theta}) = (\theta - \tilde{\theta})^2$, la cual bajo su minimización, conduce al valor esperado posteriori $E[\theta/H]$. Para exponer este valor esperado multiparamétrico, de un modo más concreto se usa al vector unitario $1_k^t = [1 \dots 1]^t$ y para $k = 14$, así se expresa como la siguiente igualdad:

$$\hat{\theta} = E[\theta/H] = [diag(Y * 1_{14})]^{-1} * Y$$

La matriz como resultado en si se detalla de manera extensa en Anexos. Por otro lado el proceso de estimación aun no ha culminado para $\hat{\theta}$, ya que esta matriz debe reajustarse con el uso de la matriz de estados simples (Sullivan & Woodall, 1994), este proceso necesita primero a la matriz que se denota como $\Pi = (\pi_1, \pi_2, \dots, \pi_k)$, la misma se estima de manera más directa que θ , debido a que solo son los parámetros de una Multinomial.

Es sencillo verificar que el estimador Máximo Verosímil del vector de parámetros de Π , considerando muestras de tamaño m , esta dado por $\hat{\Pi}_{MV} = \frac{1}{m}(m_1, m_2, \dots, m_k)$ en donde m_i es el total de observación en la i -ésima categoría, y m es el total de la muestra observada.

Ahora es necesario mejorar la estimación anterior con la incorporación de la priori de Jeffrey's, la cual tiene una distribución Dirichlet con vector de parámetros²⁶ $\Pi_0 = (\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2})$, y bajo el mismo mecanismo de trabajo que se uso en la estimación de $\hat{\theta}$, se obtiene la estimación posteriori de Π , la cual conlleva a una Dirichlet posteriori con parámetros $\Pi_{ps} = (m_1 + \frac{1}{2}, m_2 + \frac{1}{2}, \dots, m_k + \frac{1}{2})$ y esperanza:

$$\hat{\Pi} = E[\Pi/m] = \frac{1}{traza (diag (\Pi_{ps}))} * \Pi_{ps}$$

²⁶ Este paso ya se ha verificado en el capítulo II, bajo el uso de la información de Fisher.

Como solo son vectores columnas es posible ver su composición, el vector de datos M y el vector de estimación posteriori, siendo cada uno de ellos definidos como:

$$M = \begin{bmatrix} 13 \\ 12 \\ 13 \\ 12 \\ 13 \\ 12 \\ 12 \\ 13 \\ 12 \\ 13 \\ 11 \\ 13 \\ 12 \\ 12 \end{bmatrix}; \quad \hat{\Pi} = \begin{bmatrix} 0,075 \\ 0,069 \\ 0,075 \\ 0,069 \\ 0,075 \\ 0,069 \\ 0,069 \\ 0,075 \\ 0,069 \\ 0,075 \\ 0,064 \\ 0,075 \\ 0,069 \\ 0,069 \end{bmatrix}$$

Desde este punto, y con el uso de $\hat{\Pi}$ y $\hat{\theta}$, se obtiene la denominada matriz ajustada de transición (Sullivan & Woodall, 1994), se la denotara por $\hat{\theta}_{adj}$, la misma no es otra cosa que la normalización en fila para el producto de Hadamard, el cual como sabe, está determinado mediante: $(A \otimes B)_{i,j} = (A)_{i,j} * (B)_{i,j}$ pero para el caso se define $B = \hat{\theta}$ y la matriz A de dimensión $k \times k$ (que para el caso es $k = 14$) mediante²⁷:

$$A = [\hat{\Pi}^t / \hat{\Pi}^t / \dots / \hat{\Pi}^t]$$

Y la matriz ajustada de Markov $\hat{\theta}_{adj}$, se normaliza al incorporar el vector columna $1_k = (1 \ 1 \ \dots \ 1)$, (tanto la matriz no ajustada $\hat{\theta}$ y la ajustada $\hat{\theta}_{adj}$, se detallan en Anexos) con $k = 14$ determinándose finalmente como:

$$\hat{\theta}_{adj} = [diag(A * \hat{\theta} * 1_k)]^{-1} * A * \hat{\theta}$$

3.2.2. Números y Variables Aleatorias Difusas

Creada la matriz de transición de Markov, se prosigue con el cálculo de los números difusos y algunos elementos relacionados a ellos. Esta fase en sí, no se menciona en el esquema de las Series de Tiempo Difusas para el capítulo anterior, ya que no se había incorporado el manejo de números difusos en sí. El nexa necesario para poder hacer esto, es la implementación de las variables aleatorias difusas, del siguiente modo:

“El hecho de que exista una relación entre los Clusters cuantificada en la matriz $\hat{\theta}_{adj}$ que contiene distribuciones de probabilidad a nivel fila, implica que es posible esta-

²⁷ Cabe recordar que el operador “/” en vectores solo apila los vectores fila, uno debajo del otro.

blecer una variable aleatoria difusa ω_t del periodo t , condicionada al estado anterior $t - 1$, con la distribución de probabilidad $(\tilde{x}_j; \hat{\theta}_{i,j})$, en donde $\omega_{t-1} \in A_i$ ”.

En el párrafo anterior, se acepta el hecho de definir k variables aleatorias difusas (Kwakernaak, 1978), las cuales usan los parámetros estimados de la cadena de Markov, pero estas toman valores en un espacio difuso, en resumidas cuentas se sostiene que:

$$\omega_t / \omega_{t-1} \in A_i \sim \{ \tilde{x}_j; \Pr(\omega_t = \tilde{x}_j) = \hat{\theta}_{i,j} \}$$

Ahora el asunto en sí, es el de obtener dichos números en base a los clusters que se han conformado de manera exhaustiva $\{A_i\}$. Se conoce de varios métodos para aproximar la función de pertenencia de un conjunto difuso (Ross, 2004), pero dado el enfoque de manejo de números difusos, se dará prioridad a los Números Trapezoidales ya que se desarrollaron para generar mayor eficiencia en su manejo.

Cada Cluster A_i , debe ser expresado ahora por un solo número difuso. Esto se hace posible con el manejo de los estadísticos dentro de cada Cluster, acercándose de este modo a los parámetros del número. El proceso se reduce mediante en la obtención de cuatro valores, para así definir al número trapezoidal $\tilde{x}_j = (\overset{j}{1}x, \overset{j}{2}x, \overset{j}{3}x, \overset{j}{4}x)$.

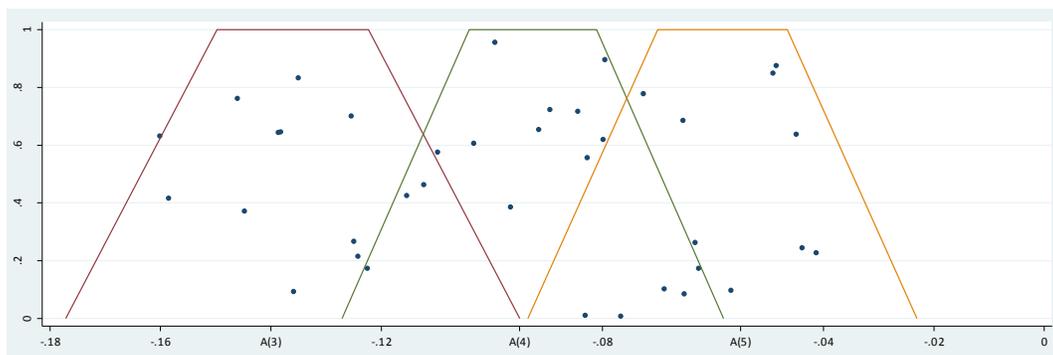


Ilustración 21. Representación de los Clusters conformados (A3, A4 & A5)

Existe una aproximación aceptable de los valores del número trapezoidal en función a datos recolectados (Klir & Yuan, 1995), esta misma usa estadísticos como la media y la desviación estándar para dicho acercamiento, y mediante la siguiente serie de ecuaciones: $\overset{j}{1}x \cong \bar{x}_j - 3 * s_j$; $\overset{j}{2}x \cong \bar{x}_j - s_j$; $\overset{j}{3}x \cong \bar{x}_j + s_j$; y $\overset{j}{4}x \cong \bar{x}_j + 3 * s_j$, se obtienen los parámetros del número trapezoidal (Ilustración 21).

En resumen, hasta este punto se obtuvo la matriz de Markov ajustada $\hat{\theta}_{adj}$ por un lado, y por otro el vector difuso denominado así, ya que se estaría compuesto por cada número trapezoidal difuso $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k]$.

TABLA 5. APROXIMACIÓN DE NÚMEROS DIFUSOS TRAPEZOIDALES

	j_1x	j_2x	j_3x	j_4x
A_1	-0,48	-0,34	-0,21	-0,07
A_2	-0,22	-0,19	-0,17	-0,14
A_3	-0,18	-0,15	-0,12	-0,09
A_4	-0,13	-0,10	-0,08	-0,06
A_5	-0,09	-0,07	-0,05	-0,02
A_6	-0,06	-0,04	-0,01	0,01
A_7	-0,02	0,00	0,01	0,02
A_8	0,00	0,02	0,04	0,06
A_9	0,03	0,04	0,06	0,07
A_{10}	0,04	0,07	0,09	0,12
A_{11}	0,09	0,10	0,12	0,14
A_{12}	0,12	0,14	0,15	0,17
A_{13}	0,09	0,17	0,25	0,34
A_{14}	0,25	0,30	0,36	0,41

3.2.3. Calculo de Pronósticos

Para el cómputo de los pronósticos, u obtención de valores pronosticados de la serie, se hace uso de dos conceptos: *número difuso* como tal, y de *valores Defuzzificados*, que ya desglosaron en el capitulo anterior.

El cálculo de los pronósticos, con el uso de las variables aleatorias difusas y por ende con los números trapezoidales, procede estableciendo la ecuación habitual de Chen: $\hat{F}_i(t) = \frac{1}{k} \sum_{h=1}^k m_{j,h}$, se incorporan los ponderadores y la ecuación se extiende de tal modo que: $\hat{F}_i(t) = \frac{1}{\sum_j w_j} \sum_{h=1}^k w_j m_{j,h}$, de manera más formal y con el uso de $\hat{\theta}_{adj}$ y \tilde{X} :

$$\begin{aligned} \hat{F}_i(t) &= \frac{1}{\sum_j \hat{\theta}_{i,j}} \sum_j \tilde{x}_j * \hat{\theta}_{i,j} \\ \hat{F}_i(t) &= \sum_j \tilde{x}_j * \hat{\theta}_{i,j} \\ \hat{F}_i(t) &= \mathbb{E}[\omega_t / \omega_{t-1} \in A_i] \end{aligned}$$

Siendo solamente la esperanza condicional en el espacio difuso $\mathbb{E}[\omega_t / \omega_{t-1} \in A_i]$, y el término $\sum_j \tilde{x}_j * \hat{\theta}_{i,j}$ ya se ha desarrollado anteriormente para el caso de números trapezoidales, como la quarteta: $[\sum_j {}^j_1x * \hat{\theta}_{i,j} ; \sum_j {}^j_2x * \hat{\theta}_{i,j} , \sum_j {}^j_3x * \hat{\theta}_{i,j} ; \sum_j {}^j_4x * \hat{\theta}_{i,j}]$

El proceso de la obtención de todas estas esperanzas $\mathbb{E}[\omega_t / \omega_{t-1} \in A_i]$ no es posible realizarlo mediante una expresión concreta, y esto se debe a que el término \tilde{X} , no es un número real en sí, sobre el cual se pueda usar operadores matriciales de manera normal. De todos modos con la ayuda de software adecuado, se obtuvo las $k = 14$ esperanzas condicionales resultaron en la Tabla 6.

Con estos valores esperados, se procede al cálculo de los pronósticos, pero es necesaria unas aclaraciones previas. La primera es que estas esperanzas son números difusos, y para poderlas conservar como tales, se debe aceptar el hecho de que cuando $a \in \mathbb{R}$ entonces también $a \in \tilde{\mathcal{F}}_0(\mathbb{R}^d)$, es decir que todo número real pertenece al espacio de los números difusos, y particularizando para el caso de número trapezoidal $\tilde{a} = (a_1, a_2, a_3, a_4)$ donde $a_i = a$.

El otro hecho es que se tomo una diferencia, y previamente la transformación logarítmica, estas operaciones se subsanan²⁸ como $\hat{Y}_t = \exp(\tilde{Z}_{t-1} + \mathbb{E}[\omega_t/\omega_{t-1} \in A_i])$. Se recalca que el término \hat{Y}_t es un número difuso, así que en la grafica que a continuación se expone, se lo expresa bajo dos de sus cuatro componentes²⁹ ($y1_f$ & $y4_f$ representan a j_1x & j_4x).

TABLA 6. VALORES ESPERADOS TRAPEZOIDALES

	\tilde{x}_j			
	j_1x	j_2x	j_3x	j_4x
A_1	0,03	0,06	0,09	0,12
A_2	0,07	0,12	0,16	0,21
A_3	0,05	0,08	0,11	0,15
A_4	0,04	0,07	0,10	0,13
A_5	0,04	0,06	0,08	0,10
A_6	0,03	0,07	0,11	0,15
A_7	-0,12	-0,07	-0,02	0,03
A_8	-0,10	-0,07	-0,04	-0,02
A_9	-0,02	0,00	0,03	0,05
A_{10}	-0,07	-0,05	-0,02	0,00
A_{11}	-0,02	0,01	0,04	0,07
A_{12}	-0,13	-0,09	-0,05	-0,01
A_{13}	-0,16	-0,11	-0,05	0,00
A_{14}	-0,28	-0,21	-0,14	-0,07

²⁸ Al igual que el eje real, los números difusos cumplen ciertas transformaciones, estas se desglosan a más detalle en Anexos.

²⁹ Graficar las cuatro componentes de manera global, disminuye considerablemente la visibilidad de los datos, a nivel individual presenta mejor manejo.

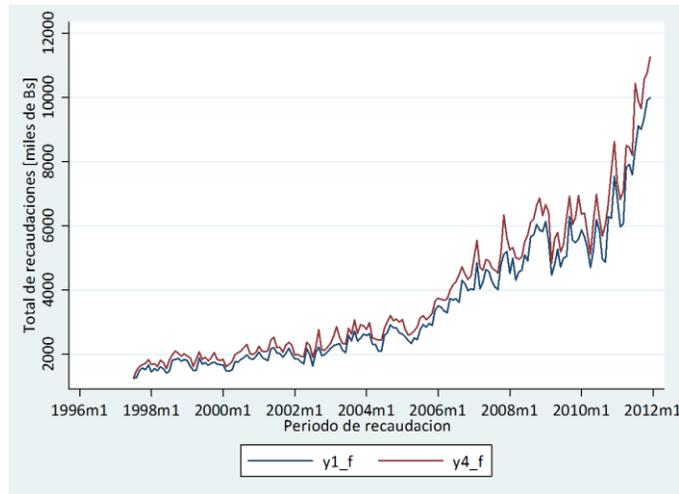


Ilustración 22. Valores pronosticados mediante números difusos

En la segunda manera se acerca al original modelo de series de tiempo difusa (Chen, 1996), ya que simplemente se deberían de usar los valores Defuzzificados de cada conjunto, tomando un rumbo alternativo, se procede a la defuzzificacion de los términos de \hat{Y}_t , mediante la técnica del centroide, se dio a entender que este método tiene directa relación con el valor esperado de una variable aleatoria en \mathbb{R} , con este hecho se hace uso del valor esperado de una distribución trapezoidal estándar³⁰ X de parámetros $(\alpha, \beta, \gamma, \delta)$:

$$E[X/\alpha, \beta, \gamma, \delta] = \frac{(\beta - \alpha)(\alpha + 2\beta) - 3(\beta^2 - \gamma^2) + (\delta - \gamma)(2\gamma + \delta)}{3(\delta + \gamma - \beta - \alpha)}$$

Con el uso de esta ecuación se procedió a Defuzzificar los datos \hat{Y}_t , y compararlos con la serie original, esto de manera grafica.

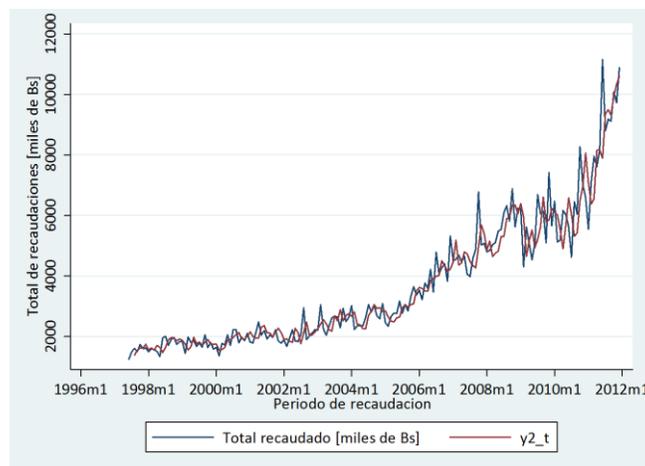


Ilustración 23. Serie pronosticada defuzzificada y serie real

³⁰ El desglose de esta distribución, y sus propiedades más relevantes se desglosan en los acápites posteriores.

3.2.4. Medidas de Bondad de Ajuste

Entre las medidas usadas en este acápite se opta por la de mayor uso, es decir la del Error Cuadrático Medio (SME por sus siglas en ingles), la misma se precisa mediante la ecuación $MSE = \frac{1}{T} \sum_{t=1}^T (Y_t - \hat{Y}_t)^2$.

El aplicar la anterior ecuación está limitado a que los valores reales Y_t y los valores pronosticados o estimados \hat{Y}_t deben de pertenecer a espacio real \mathbb{R} . Para resolver esto, ya se describió en el cálculo de los pronósticos que se optó en segunda instancia en la defuzzificación, como se ve en la Ilustración 23.

Dado que se usa un método de defuzzificación, que es el del *centroide* concretamente, es vio pertinente compararlo con otro método de su misma índole. El método que es adecuado para el caso de números trapezoidales, es el de *Promedio de Máximos*, método de Defuzzificación que se resume en el uso del soporte del número difuso trapezoidal, siendo el promedio simple: $x_0 = \frac{1}{2}(x_2 + x_3)$.

Tanto el método del centroide, como el de promedio de máximos, resultan en errores cuadráticos medios de: 33.6184 y 33.4599 respectivamente. La regresión difusa, que es una técnica en base a números difusos, maneja además la medida de bondad de ajuste que la denomina “*R-cuadrado híbrido*”, esta medida es mucho más elaborada, ya que se basa en distancias difusas estandarizadas y otros conceptos más extensos, con lo cual no se la considera ya que no permitiría una comparación con el siguiente método.

3.3. Metodología BOX & JENKINS: Modelo ARIMA

Con fines comparativos, y sin entrar en mucho detalle, se procedió a la implementación de la metodología de Box & Jenkins, con la cual se identificó que poseía parámetros para un *ARIMA* ($p = 0, d = 1, q = 1$) *SARIMA* ($P = 1, D = 1, Q = 0$). Es importante conocer las estimaciones y las pruebas asociadas al modelo *ARIMA*, con lo cual se dedica un apartado entero en los apéndices.

Tras la estimación de parámetros del modelo *ARIMA* (0,1,1)*SARIMA* ($P = 1, D = 1, Q = 0$) se derivó a la obtención de los valores estimado de la serie original, los mismos conllevaron a la obtención de un $ECM = 290407,1$. A continuación se muestran de manera paralela con los datos originales en la Ilustración 24.

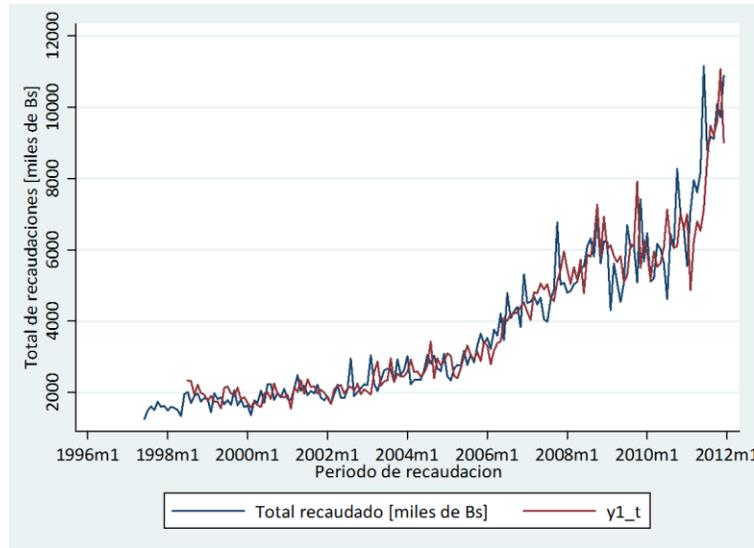


Ilustración 24. Serie estimada ARIMA (0;1;1) SARIMA(1,1,0) y serie real

3.4. Comparación entre Modelo de Markov Difuso y Arima

Bajo el uso de las medidas de bondad de ajuste, y particularmente con el error cuadrático medio, se hace el cotejo de estos dos métodos, el Modelo de Markov Difuso y el Modelo ARIMA. La tabla siguiente hace uso de los ECM, para los ambos métodos, pero en el caso del Modelo de Markov, contempla a los dos métodos de defuzzificación, el promedio de máximos y el del centroide.

TABLA 7. COMPARACION DE ECM ENTRE MODELOS

Modelo	ECM
ARIMA(0,1,1) SARIMA(1,1,0)	290407,1
Markov (k=14): Centroide	336184,4
Markov (k=14): Promedio de Máximos	334599,8

Entre los métodos de defuzzificación existe una diferencia algo tenue, pero al ser cotejados con el modelo ARIMA la diferencia es considerable, dando a conocer que ambos métodos poseen un ajuste menor. Si bien el número de cluster influye en el proceso de estimación, se procede a variar este parámetro con fines comparativos, desde un $k = 10$ a $k = 80$, tomando estos valores a prueba (Arutchelvan, Sivasta, & Jaganathan, 2010).

La grafica que a continuación se muestra nos da un panorama de cómo evoluciona el ajuste del Modelo de Markov Difuso en términos de su *ECM*, esto conforme se incrementa el número de clusters implementados, y de manera paralela se expone los valores obtenidos en un inicio, los del modelo de $k=14$, y los del ARIMA SARIMA respectivamente. Notándose una superioridad del modelo ARIMA-SARIMA, hasta un número de $k=30$, de ahí en adelante, el modelo de Markov, brinda ECM menores.

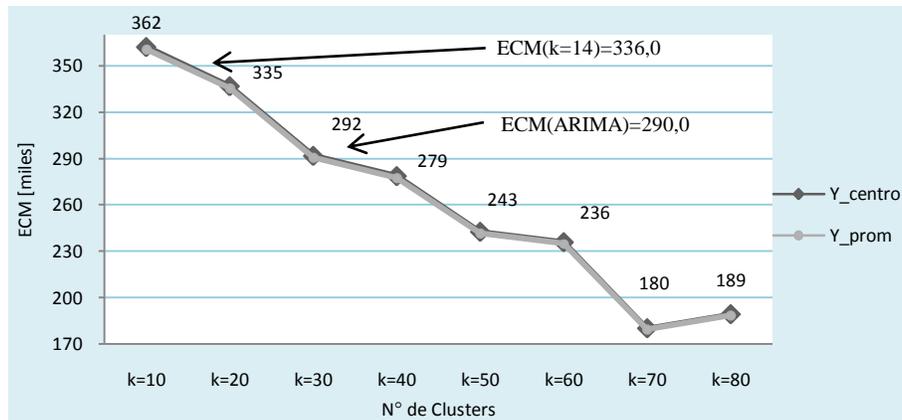


Ilustración 25. Evolución de ECM para diferentes números de Clusters

3.5. Conclusiones

A nivel global se rescata una conclusión muy importante, la misma es utilidad de la teoría difusa, tanto en los números, como en las variables aleatorias asociadas a este campo, ya que brindan para el caso de series de tiempo difusas, un desempeño menor inicialmente, pero que mejora conforme incrementan el número de clusters, esto en términos del manejo de ajuste a través del ECM.

Es prudente mencionar que con el número de catorce clusters³¹, el cual inicialmente se manejaba no se puede alcanzar un ajuste semejante al modelo ARIMA, esto para la serie en particular, siendo esto subsanado.

3.5.1. Respecto al Modelo de Markov Difuso

El modelo de Markov Difuso, si bien estaba desfavorecido en un inicio, evoluciona propiciamente su ajuste, y es meritorio reconocer el respaldo estadístico que maneja durante toda su construcción, estimación e incorporación de métodos bayesianos.

Por otro lado los procesos de Markov, se basan principalmente en el manejo de matrices de transición, el aporte es la metodología desglosada en el proceso de estimación y cálculo para estas matrices, todo esto en base a datos observados, más aun con la incorporación de mejoras a través del *Método Bayesiano* y el uso de la *Distribución Dirichlet*.

Finalmente las variables aleatorias, denominadas difusas, que se asocian las matrices antes mencionadas, son un progreso considerable de la estadística, ya que se definen sobre campos que no son necesariamente el conjunto de números reales, brindando de esta manera el manejo y mutuo actuar bajo conceptos probabilísticos y difusos.

³¹ El número recomendable de clusters a ser manejados, se expuso en el capítulo anterior, el mismo se apega es: $k = \sqrt{T}$.

3.5.2. Respecto a los Resultados

El modelo de Markov Difuso, al ser aplicado en la estimación de la serie: “*Total Recaudado [en miles de Bs]*” para las Administradoras de Fondos de Pensiones en el departamento de Potosí, inicialmente se observa resultados desfavorables en comparación con un modelo ARIMA, sin embargo se muestra que el incremento del número de clusters disminuye sus ECM.

Recordando que se trata de un caso aislado para series de datos temporales, y con los resultados obtenidos no se rechaza la hipótesis de investigación: *es posible generar una metodología alternativa de pronósticos de series de tiempo univariado en base al Modelo de Markov y la Teoría Difusa.*

3.6. Recomendaciones

En primer lugar se recomienda que la aplicación de este modelo es de elección del investigador o usuario, ya que el mismo modelo aun está sujeto a mejoras y avances, y conforme se avanza la elaboración de toda la investigación de los mismos se vio pertinente rescatar algunas sugerencias e incorporaciones tentativas para ser estudiadas, estas son: *Análisis de tendencias y Estacionariedad*, (Bernd Möller, 2006), *Incorporación del Modelo Multivariante* (Bulut & Shigeru, 2012), *Modelos FARIMA* por presentar componentes difusas en sus parámetros (Tseng & Tzeng, 2001).

Como el segundo tema relevante de la investigación, es la “*Teoría de las Cadenas de Markov Ocultas*”, que lleva a un nivel más avanzado al clásica teoría de los Procesos de Markov Discretos, la misma presenta aun varios avances recientes.

Y por último, vale la pena mencionar algunos modelos como asociados a la teoría difusa, estos son: *Regresión Posibilista*, *Regresión Difusa*, entre otros ya que comparten semejanza a los modelos logísticos, pero los primeros son de naturaleza borrosa, y su manejo encaja en otro campo diferente de los reales y de la probabilidad.

Anexos

A.1. Sigma algebra & Teoría de la Probabilidad

Para demarcar los conceptos en el campo difuso, los principios que se siguen en probabilidad sobre números reales son los de *sigma algebra*, y buscando enfatizar en estos antes mencionados, se aclaran las siguientes definiciones, y sea X un conjunto arbitrario:

Definición A.1. *Un σ -algebra \mathcal{F} de subconjuntos de X , es una colección de subconjuntos de X que satisface las siguientes condiciones:*

- i. *El conjunto $\emptyset \in \mathcal{F}$.*
- ii. *Si $B \in \mathcal{F}$ entonces el complemento $\bar{B} \in \mathcal{F}$*
- iii. *Si E_1, E_2, \dots es una colección contable de conjuntos en \mathcal{F} , entonces la unión $\bigcup_{i=1}^{\infty} E_i$ también pertenece a \mathcal{F} .*

En ocasiones se denomina como “*sigma-algebra*”, en vez de “*sigma-algebra de subconjuntos de X* ”, algunos ejemplos convencionales de sigma-algebra son:

- El conjunto de conjuntos $\{\emptyset, X\}$ es un sigma algebra de X .
- El conjunto $\mathcal{P}(X)$ de todos los subconjuntos de X es un sigma algebra.

Estos dos casos acotan que cualquier conjunto de sigmas algebras \mathcal{F} de X , deberían de estar entre los mismos, es decir:

$$\{\emptyset, X\} \subset \mathcal{F} \subset \mathcal{P}(X)$$

Proposición I. *Sea \mathcal{F} un sigma algebra de subconjuntos de X .*

- i. *$X \in \mathcal{F}$.*
- ii. *Si $A_1, A_2, \dots, A_n \in \mathcal{F}$, entonces $\bigcup_{i=1}^n A_i \in \mathcal{F}$.*
- iii. *Si $A_1, A_2, \dots, A_n \in \mathcal{F}$, entonces $\bigcap_{i=1}^n A_i \in \mathcal{F}$.*
- iv. *Si A_1, A_2, \dots es una colección contable de conjuntos en \mathcal{F} entonces $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.*
- v. *Si $A, B \in \mathcal{F}$ entonces $A - B \in \mathcal{F}$.*

Prueba.-

- Como $\emptyset \in \mathcal{F}$, y $\emptyset^c = X$ entonces $X \in \mathcal{F}$, comprobándose (i).
- Es posible escribir como:

$$A_1 \cup \dots \cup A_n = A_1 \cup \dots \cup A_n \cup \emptyset \cup \dots$$

Así este último pertenece por definición a \mathcal{F} , comprobándose (ii).

- Debido a las leyes de Morgan, las siguientes dos expresiones son equivalentes:

$$A_1 \cap \dots \cap A_n = (A_1^c \cup \dots \cup A_n^c)^c$$

Como cada conjunto $A_i^c \in \mathcal{F}$ y como $A_1^c \cup \dots \cup A_n^c \in \mathcal{F}$ por (i), entonces su complemento también pertenece $(A_1^c \cup \dots \cup A_n^c)^c \in \mathcal{F}$, comprobándose (iii).

- De manera similar se desglosa el término $\bigcap_{i=1}^{\infty} A_i = [\bigcup_{i=1}^{\infty} A_i^c]^c$, ya que $A_i^c \in \mathcal{F}$ y como el conjunto es cerrado respecto al complemento, comprobándose (iv).
- Finalmente $A - B = A \cap B^c$, usando el hecho de (iii) se verifica de manera inmediata, comprobándose (v).

Sea X un conjunto y \mathcal{B} una colección no vacía de subconjuntos de X . El “*más pequeño*” de los algebra de todos los conjuntos de \mathcal{B} es denotado como: $\sigma(\mathcal{B})$ y es llamado el “ σ -algebra generador de \mathcal{B} ”.

El término más pequeño aquí significa que cualquier σ -algebra que contiene elementos de \mathcal{B} , tendría que contener a todos los conjuntos de $\sigma(\mathcal{B})$, dicho de otro modo:

Proposición II. Si G es una colección de sigma algebras de subconjuntos de X entonces la intersección $\bigcap G$ es denominada una sigma algebra de conjuntos de X :

$$\bigcap G = \{A \subset X \mid A \in \mathcal{F} \text{ para cada } \mathcal{F} \in G\}$$

Es entonces $\bigcap G$ el que consiste en todos los conjuntos A que pertenecen a cada sigma-algebra \mathcal{F} de G . Las definiciones y proposiciones antes mencionadas dan un panorama lo bastante sólido, con lo cual ahora se abarca ahora los conceptos de la *teoría de la medida*.

Definición A.2. Sea X un conjunto no vacío y sea \mathcal{F} un σ -algebra sobre X , una función μ con dominio en σ -algebra y valores entre $[0, +\infty)$ se dice que es contable aditiva si satisface:

$$\mu(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$$

Donde $\{A_i\}$ es una secuencia infinita de conjuntos disjuntos de \mathcal{F} que pertenecen a \mathcal{F} , y finalmente una función $\mu: \mathcal{F} \rightarrow [0, +\infty)$ que satisface $\mu(\emptyset) = 0$ y que es contable aditiva, se denomina una medida sobre \mathcal{F} .

Definición A.3. Sea X un conjunto no vacío, sea \mathcal{F} un σ -algebra sobre X , y sea una medida μ sobre \mathcal{F} , entonces la tripleta (X, \mathcal{F}, μ) es denominada espacio de medida. Igualmente, si X es un conjunto no vacío y \mathcal{F} un σ -algebra sobre X , entonces el par (X, \mathcal{F}) se denomina espacio medible.

Si bien es posible establecer varios teoremas adicionales a estas definiciones, se ve mas efectivamente avanzar al campo de la probabilidad.

Definición A.4. Sea X un conjunto no vacío, sea \mathcal{F} un σ -álgebra sobre X , una función de probabilidad P con dominio en \mathcal{F} , satisface las siguientes condiciones.

- i. $P(A) \geq 0$ para todo A en \mathcal{F} .
- ii. $P(X) = 1$.
- iii. Para cada secuencia infinita $\{A_i\}$ de conjuntos disjuntos que pertenecen a \mathcal{F} , se tiene que:

$$P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$$

La tripleta (X, \mathcal{F}, P) que a su vez es un espacio de medida, es denominada de manera más específica como un *espacio de probabilidad* y la medida P es denominada *medida de probabilidad*.

Las tres propiedades dadas en la definición A.4 se refieren generalmente como los axiomas de Probabilidad (o los axiomas Kolmogorov). Cualquier función P que satisface los axiomas Probabilidad de que se llama una función de probabilidad. La definición axiomática no hace tratar de contar lo particular función P para elegir, sino que se limita a exigir P para satisfacer la axiomas.

A.1.1. Nociones sobre Variables Aleatorias Difusas

La noción de variables aleatorias difusas se introdujo por primera vez (Kwakernaak, 1978) de la siguiente manera:

Sea la tripleta $(\Omega, \mathfrak{F}, \mathcal{P})$ un espacio de probabilidad, suponga que U es variable aleatoria definida sobre esta tripleta. Asumiendo ahora que se percibe esta variable aleatoria U a través de conjuntos de ventanas $W_i, i \in J$, con J un conjunto finito o numerable, cada uno representando a un intervalo de la recta real, de tal modo que $W_i \cap W_j = \emptyset$ para $i \neq j$, y $\bigcup_{i \in J} W_i = \mathbb{R}$. Percibir la variable aleatoria a través de estas W_i significa que para cada w solo se puede establecer que $U(\omega) \in W_i$ para algún $i \in J$.

Definiendo la función $I_i: \mathbb{R} \rightarrow [0,1]$ como la función característica del conjunto W_i . Además sea S el espacio de todas las funciones continuas por tramos de asignación $\mathbb{R} \rightarrow [0,1]$. A continuación, definir la percepción de la variable aleatoria U , tal como se describió anteriormente, como el mapeo $X: \Omega \rightarrow S$ dada por: $\omega \rightarrow X_\omega$.

Con $X_\omega = I_i$ si y solo si $U(\omega) \in W_i$, esto significa que asociamos con cada $\omega \in \Omega$ no un número real $U(\omega)$, como en el caso de una variable aleatoria ordinaria, pero si una función característica X_ω que es un elemento de S .

El mapeo $X: \Omega \rightarrow S$ descrito anteriormente caracteriza un tipo especial de variable aleatoria difusa. La variable aleatoria U , del cual esta variable aleatoria difusa es una percepción, se llama variable aleatoria difusa original.

Definición A.4. Sea el mapeo $\xi: \Omega \rightarrow \mathfrak{F}$, donde \mathfrak{F} es el conjunto de todos los números difusos, denotamos la imagen de $\omega \in \Omega$ mediante $\xi(\omega) = (\mathbb{R}, X_\omega, a_\omega)$, con $X_\omega \in \mathcal{S}$ y $a_\omega: \mathbb{R} \rightarrow \mathcal{P}$, y el mapeo $X: \Omega \rightarrow \mathcal{S}$ especificado como:

$$\omega \xrightarrow{X} X_\omega$$

Se requiere además que para cada $\mu \in (0,1]$ ambos U_μ^* y U_μ^{**} definidos por:

$$\begin{aligned} U_\mu^*(\omega) &= \inf\{x \in \mathbb{R} / X_\omega(x) \geq \mu\} \\ U_\mu^{**}(\omega) &= \sup\{x \in \mathbb{R} / X_\omega(x) \geq \mu\} \end{aligned}$$

Sean variables aleatorias reales valoradas, definidas sobre $(\Omega, \mathfrak{F}, \mathcal{P})$ que satisfacen:

$$(\forall \omega \in \Omega), \quad X_\omega(U_\mu^*(\omega)) \geq \mu, \quad X_\omega(U_\mu^{**}(\omega)) \geq \mu.$$

Y finalmente, para cada $\omega \in \Omega$ y cada $x \in \mathbb{R}$, $a_\omega(x)$ es un estado:

$$a_\omega(x) = (\text{asume el valor original de "x" en el punto } \omega)$$

A.2. Números Difusos

Antes de mencionar las operaciones aritméticas de manera mas detallada es necesario exponer el principio de extensión (Zadeh, 1965). Se trata del principio fundamental en que se sustentan aquellas aplicaciones de los conjuntos difusos basados en la amplificación de conceptos matemáticos no difusos (Liu & Pedrycz, 2009).

Definición A.5. (Principio de Extensión).- Sea una función $f: X \rightarrow Y$ y si definimos a X sobre un conjunto difuso A y sea $y = f(x)$ sobre otro conjunto difuso B cuya función de pertenencia está dada por:

$$\mu_B(y) = \begin{cases} 0 & \text{si } f^{-1}(y) = \emptyset \\ \bigvee_{y=f(x)} \mu_A(x) & \text{si } f^{-1}(y) \neq \emptyset \end{cases}$$

Al desarrollar las operaciones aritméticas de números difusos, se maneja de manera muy natural la generalización del anterior principio.

Definición A.6. (Principio de Extensión Generalizado).- Sea una función $f: X_1 \times X_2 \times \dots \times X_n \rightarrow Y$, definimos a X_1, X_2, \dots, X_n sobre los conjuntos difusos A_1, A_2, \dots, A_n y sea $y = f(x_1, x_2, \dots, x_n)$ sobre otro conjunto difuso B cuya función de pertenencia está dada por:

$$\mu_B(y) = \begin{cases} 0 & \text{si } f^{-1}(y) = \emptyset \\ \bigvee_{y=f(x_1, x_2, \dots, x_n)} \bigwedge_{i=1}^n \mu_{A_i}(x_i) & \text{si } f^{-1}(y) \neq \emptyset \end{cases}$$

Algunos numero difusos como los triangulares, son casos particulares de los números LF (left-rigth), estos últimos se descomponen como un par de funciones, los cuales a su vez se expresan en alfa cortes: $A_\alpha = [A_\alpha^1, A_\alpha^2]$ siendo α el nivel de corte, A_α^1 es la función left, y A_α^2 la función righth.

A.2.1. Operadores Aritméticos

A continuación se detallan una las “Operaciones Aritméticas Monoarias” para los números difusos de tipo left-rigth, que se denota como $A_\alpha = [A_\alpha^1, A_\alpha^2]$.

Opuesto: Sea el número difuso A , el opuesto $B = -A$, viene dado como:

$$B_\alpha = [B_\alpha^1, B_\alpha^2] = [-A_\alpha^2, -A_\alpha^1]$$

Multiplicación por un Escalar: Sea el numero difuso A , y sea un escalar $k \in \mathbb{R}$, el resultado de $B = k * A$ viene dado como:

$$B_\alpha = [B_\alpha^1, B_\alpha^2] = \begin{cases} [kA_\alpha^1, kA_\alpha^2] & k > 0 \\ [kA_\alpha^2, kA_\alpha^1] & k < 0 \end{cases}$$

En caso de que $k=0$, $B_\alpha = [0,0]$ y solo $\mu_B(x) = 1$ si $x = 1$.

Inverso: Sea el número difuso A , con $Sop(A) \subseteq \mathbb{R}^+$, el número difuso inverso $B = A^{-1}$ vendrá dado como:

$$B_\alpha = [B_\alpha^1, B_\alpha^2] = [1/A_\alpha^2, 1/A_\alpha^1]$$

Logaritmo: Sea el número difuso A , con $Sop(A) \subseteq \mathbb{R}^+$, el número difuso que representaría su logaritmo neperiano $B = \ln(A)$ vendrá dado como:

$$B_\alpha = [B_\alpha^1, B_\alpha^2] = [\ln(A_\alpha^1), \ln(A_\alpha^2)]$$

Exponencial: Sea el número difuso A , el número difuso que representaría su exponencial $B = \exp(A)$ vendrá dado como:

$$B_\alpha = [B_\alpha^1, B_\alpha^2] = [\exp(A_\alpha^1), \exp(A_\alpha^2)]$$

Potencia: Sea el número difuso A , con $Sop(A) \subseteq \mathbb{R}^+$, y sea un escalar $k \in \mathbb{R} - \{0\}$ el número difuso que representaría su potencia $B = A^k$ vendrá dado como:

$$B_\alpha = [B_\alpha^1, B_\alpha^2] = \begin{cases} [(A_\alpha^1)^k, (A_\alpha^2)^k] & k > 0 \\ [(A_\alpha^2)^k, (A_\alpha^1)^k] & k < 0 \end{cases}$$

En caso de que $k=0$, $B_\alpha = [1,1]$ y solo $\mu_B(x) = 1$ si $x = 1$.

Para dos números difusos denotados como $A_\alpha = [A_\alpha^1, A_\alpha^2]$ y $B_\alpha = [B_\alpha^1, B_\alpha^2]$. A continuación se detallan las “Operaciones Aritméticas Binarias”.

Suma de Números Difusos: Para dos números difusos A y B la suma $C = A + B$ viene dada por:

$$C_\alpha = [C_\alpha^1, C_\alpha^2] = [A_\alpha^1 + B_\alpha^1, A_\alpha^2 + B_\alpha^2]$$

Resta de Números Difusos: Para dos números difusos A y B la diferencia entre ellos $C = A - B$ viene dada por:

$$C_\alpha = [C_\alpha^1, C_\alpha^2] = [A_\alpha^1 - B_\alpha^2, A_\alpha^2 - B_\alpha^1]$$

Producto de Números Difusos: Para dos números difusos A y B la diferencia entre ellos $C = A * B$ viene dada por:

$$C_\alpha = [C_\alpha^1, C_\alpha^2] = [A_\alpha^1 * B_\alpha^1, A_\alpha^2 * B_\alpha^2]$$

Lo anterior se cumple siempre y cuando $Sop(A), Sop(B) \subseteq \mathbb{R}^+$, caso contrario la expresión general está dada por:

$$C_\alpha = [C_\alpha^1, C_\alpha^2] = \left[\begin{array}{l} \text{Min}\{A_\alpha^1 * B_\alpha^1, A_\alpha^1 * B_\alpha^2, A_\alpha^2 * B_\alpha^1, A_\alpha^2 * B_\alpha^2\}, \\ \text{Max}\{A_\alpha^1 * B_\alpha^1, A_\alpha^1 * B_\alpha^2, A_\alpha^2 * B_\alpha^1, A_\alpha^2 * B_\alpha^2\} \end{array} \right]$$

División de Números Difusos: Para dos números difusos A y B la diferencia entre ellos $C = A \div B$ viene dada por:

$$C_\alpha = [C_\alpha^1, C_\alpha^2] = [A_\alpha^1/B_\alpha^2, A_\alpha^2/B_\alpha^1]$$

Lo anterior se cumple siempre y cuando $Sop(A), Sop(B) \subseteq \mathbb{R}^+$, caso contrario la expresión general está dada por:

$$C_\alpha = [C_\alpha^1, C_\alpha^2] = \left[\begin{array}{l} \text{Min}\{A_\alpha^1/B_\alpha^1, A_\alpha^2/B_\alpha^1, A_\alpha^1/B_\alpha^2, A_\alpha^2/B_\alpha^2\}, \\ \text{Max}\{A_\alpha^1/B_\alpha^1, A_\alpha^2/B_\alpha^1, A_\alpha^1/B_\alpha^2, A_\alpha^2/B_\alpha^2\} \end{array} \right]$$

Una particularización son los *números trapezoidales*, de los cuales se detalla sus operadores aritméticos. Y dado un número trapezoidales $A = [a_1, a_2, a_3, a_4]$

Opuesto: Definido como $B = -A$

$$B = [b_1, b_2, b_3, b_4] = [-a_4, -a_3, -a_2, -a_1]$$

Multiplicación por un escalar: Definido como $B = k * A$

$$B = [b_1, b_2, b_3, b_4] = \begin{cases} [ka_4, ka_3, ka_2, ka_1] & k < 0 \\ [ka_1, ka_2, ka_3, ka_4] & k > 0 \end{cases}$$

Pseudo-Inverso: Definido como $B = A^{-1}$, el numero B no es exactamente trapezoidal, pero un buena aproximación al mismo se asegura si $\{0 \notin sop(A)\}$.

$$B = [b_1, b_2, b_3, b_4] \approx \left[\frac{1}{a_4}, \frac{1}{a_3}, \frac{1}{a_2}, \frac{1}{a_1} \right]$$

Suma: Definido como $C = A + B$.

$$C = [c_1, c_2, c_3, c_4] = [a_1 + b_1, a_2 + b_2, a_3 + b_3, a_4 + b_4]$$

Resta: Definido como $C = A - B$.

$$C = [c_1, c_2, c_3, c_4] = [a_1 - b_4, a_2 - b_3, a_3 - b_2, a_4 - b_1]$$

Multiplicación: Definido como $C = A \otimes B$, el número resultante C no es exactamente trapezoidal, una buena aproximación se asegura si $Sop(A), Sop(B) \subseteq \mathbb{R}^+$, siendo como resultado:

$$C = [c_1, c_2, c_3, c_4] \approx [a_1 * b_1, a_2 * b_2, a_3 * b_3, a_4 * b_4]$$

División: Definido como $C = A \div B$, el número resultante C no es exactamente trapezoidal, una buena aproximación se asegura si $Sop(A), Sop(B) \subseteq \mathbb{R}^+$, siendo como resultado:

$$C = [c_1, c_2, c_3, c_4] \approx \left[\frac{a_1}{b_4}, \frac{a_2}{b_3}, \frac{a_3}{b_2}, \frac{a_4}{b_1} \right]$$

Tanto la multiplicación como la división genera números no trapezoidales, la verdadera función de pertenencia de C esta dada por:

$$\mu_c(x) = \begin{cases} \frac{-a_1+x*b_4}{(a_2-a_1)+x*(b_4-b_3)} & \frac{a_1}{b_4} < x \leq \frac{a_2}{b_3} \\ 1 & \frac{a_2}{b_3} < x \leq \frac{a_3}{b_2} \\ \frac{+a_4-x*b_1}{(a_4-a_3)+x*(b_2-b_1)} & \frac{a_3}{b_2} < x < \frac{a_4}{b_1} \\ 0 & \text{en otro caso} \end{cases}$$

A.2.2. Distancia entre Números Difusos

Una función $\mathcal{D}: \mathcal{C} \times \mathcal{C} \rightarrow \mathbb{R}^+$ denotada como con $X, Y, Z \in \mathcal{C}$ es denominada distancia como tal, si y solo si:

- $\mathcal{D}(X, Y) \geq 0 \quad \forall X, Y \in \mathcal{C}$
- Si $X = Y \quad \Rightarrow \quad \mathcal{D}(X, Y) = 0$
- $\mathcal{D}(X, Y) = \mathcal{D}(Y, X) \quad \forall X, Y \in \mathcal{C}$
- $\mathcal{D}(X, Z) \leq \mathcal{D}(X, Y) \oplus \mathcal{D}(Y, Z) \quad \forall X, Y, Z \in \mathcal{C}$

Donde el operador \oplus esta asociado al concepto de distancia. Ahora bien, para el caso de dos números difusos A y B ; donde (a_1, b_1) y (a_2, b_2) son los valores mínimos y máximos respectivamente, Kaufman define la distancia " Δ " como:

$$\Delta(A, B) = \Delta_L(A, B) + \Delta_R(A, B)$$

Y los términos Δ_L y Δ_R están dados por: $\Delta_L = |a_1 - b_1|$ y $\Delta_R = |a_2 - b_2|$, definida de esta manera la función $\Delta(\cdot)$ cumple las propiedades de distancia.

A.3. Distribuciones y Propiedades

En la misma continuidad, se exponen propiedades importantes sobre las densidades de probabilidad usadas en los capítulos 2 y 3.

A.3.1. Distribución Dirichlet

Definición A.7. Sea un vector aleatorio (p_1, \dots, p_k) , se dice que sigue una distribución Dirichlet con parámetros $(\alpha_1, \dots, \alpha_k)$ denotado por: $(p_1, \dots, p_k) \sim \mathcal{D}i(\alpha_1, \dots, \alpha_k)$ si satisface las siguientes condiciones:

- i. Todos los $p_i \geq 0$ y $p_1 + \dots + p_k = 1$.
- ii. $f(p_1, \dots, p_k) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} [\prod_{i=1}^{k-1} p_i^{\alpha_i - 1}] [1 - \sum_{i=1}^{k-1} p_i]^{\alpha_k - 1}$

Aclarando además algunos puntos:

- Debido a la relación funcional entre las k variables (suma igual a 1), la función de distribución conjunta es degenerada. Esto por esto que la densidad propuesta solo sobre las primeras $k - 1$ variables, y el ultimo esta dado como $p_k = 1 - \sum_{i=1}^{k-1} p_i$.
- Sin embargo, todavía hay una simetría completa entre las k parejas $(p_i; \alpha_i)$ y la densidad podría haber sido de cualquiera de los k subconjuntos de tamaño $k - 1$.
- Todos los valores de α_i son positivos.
- Cuando $k = 2$, se tiene el caso donde $p_1 \sim \mathcal{B}e(\alpha_1, \alpha_2)$, siendo este caso la generalización de las familias de distribución beta.
- Si todos los $\alpha_i = 1$ se tiene una distribución uniforme sobre el espacio simplex en \mathbb{R}^k .

Es posible realizar una sub-parametrización de esta distribución, esto para una proximidad con la ya conocida Multinomial:

$$(\alpha_1, \dots, \alpha_k) \Leftrightarrow (A, \pi_1, \dots, \pi_k)$$

Donde cada $A = \sum_i \alpha_i$ y cada uno de los π_i esta dado por: $\pi_i = \frac{\alpha_i}{A}$. Bajo la nueva notación recomendada, las **distribuciones marginales** se obtienen como:

$$\begin{aligned} f(p_i) &= \int_0^1 \dots \int_0^{1 - \sum_{j=1}^{k-1} p_j} f(p_1, \dots, p_{k-1}) dp_1 \dots dp_{k-1} \\ &= \int_0^1 \dots \int_0^{1 - \sum_{j=1}^{k-1} p_j} \frac{\Gamma(A)}{\prod_{j=1}^k \Gamma(\alpha_j)} [\prod_{j=1}^{k-1} p_j^{\alpha_j - 1}] [1 - \sum_{j=1}^{k-1} p_j]^{\alpha_k - 1} dp_{k-1} \dots dp_i \dots dp_1 \\ &= C * p_i^{\alpha_i - 1} \int_0^1 \dots \int_0^{1 - \sum_{j=1}^{k-1} p_j} [\prod_{j \neq i} p_j^{\alpha_j - 1}] [1 - \sum_{j=1}^{k-1} p_j]^{\alpha_k - 1} dp_{k-1} \dots dp_i \dots dp_1 \end{aligned}$$

Tomando previamente a $C = \frac{\Gamma(A)}{\prod_{j=1}^k \Gamma(\alpha_j)}$, se obtendrá la integral de p_{k-1} , es decir:

$$\int_0^{1-\sum_{j=1}^{k-2} p_j} p_{k-1}^{\alpha_{k-1}-1} [1 - \sum_{j=1}^{k-2} p_j - p_{k-1}]^{\alpha_{k-1}-1} dp_{k-1}$$

Un cambio de variable mediante $p_{k-1} = z * (1 - \sum_{j=1}^{k-2} p_j)$, redefine la integral como:

$$\begin{aligned} & \int_0^1 [z * (1 - \sum_{j=1}^{k-2} p_j)]^{\alpha_{k-1}-1} [1 - \sum_{j=1}^{k-2} p_j - z * (1 - \sum_{j=1}^{k-2} p_j)]^{\alpha_{k-1}-1} (1 - \sum_{j=1}^{k-2} p_j) dz \\ &= \int_0^1 [1 - \sum_{j=1}^{k-2} p_j]^{\alpha_{k-1}-1+\alpha_{k-1}-1} z^{\alpha_{k-1}-1} (1-z)^{\alpha_{k-1}-1} dz \\ &= (1 - \sum_{j=1}^{k-2} p_j)^{\alpha_{k-1}+\alpha_{k-1}-1} \int_0^1 z^{\alpha_{k-1}-1} (1-z)^{\alpha_{k-1}-1} dz \\ &= (1 - \sum_{j=1}^{k-2} p_j)^{\alpha_{k-1}+\alpha_{k-1}-1} \frac{\Gamma(\alpha_{k-1})\Gamma(\alpha_{k-1})}{\Gamma(\alpha_{k-1}+\alpha_{k-1})} \end{aligned}$$

De manera análoga se consigue las integrales posteriores p_{k-2} , p_{k-3} , y demás llegando finalmente a:

$$f(p_i) = \frac{\Gamma(A)}{\Gamma(\alpha_i)\Gamma(A-\alpha_i)} p_i^{\alpha_i-1} * (1-p_i)^{A-\alpha_i-1}$$

Siendo esta una distribución Beta legitima con parámetros $p_i \sim Be(\alpha_i; \sum_{j \neq i} \alpha_j)$, dado este hecho no es necesario desarrollar detalladamente el cálculo de sus **momentos**, y los mismos son:

$$\begin{aligned} E[p_i] &= \frac{\alpha_i}{A} \\ V[p_i] &= \frac{\alpha_i(A-\alpha_i)}{A^2(1+A)} \end{aligned}$$

Como son distribuciones Beta con parámetros $p_i \sim Be(\alpha_i; A - \alpha_i)$, la distribución de $p_i + p_j$, es de la misma manera es una Beta pero con parámetros $p_i + p_j \sim Be(\alpha_i + \alpha_j; A - \alpha_i - \alpha_j)$, así que la varianza es $V[p_i + p_j] = (\alpha_i + \alpha_j)(A - \alpha_i - \alpha_j)/A^2(1 + A)$, y al ser una suma de variables aleatorias, se deduce fácilmente que:

$$\begin{aligned} V[p_i + p_j] &= V[p_i] + V[p_j] + 2 Cov[p_i; p_j] \\ \frac{(\alpha_i + \alpha_j)(A - \alpha_i - \alpha_j)}{A^2(1 + A)} &= \frac{\alpha_i(A - \alpha_i)}{A^2(1 + A)} + \frac{\alpha_j(A - \alpha_j)}{A^2(1 + A)} + 2 Cov[p_i; p_j] \end{aligned}$$

Y al despejar así la covarianza resulta: $Cov[p_i; p_j] = -\frac{\alpha_i \alpha_j}{A^2(1+A)}$, y una consecuencia directa de esto es el coeficiente de correlación:

$$Corr[p_i; p_j] = -\sqrt{\frac{\alpha_i \alpha_j}{(A-\alpha_i)(A-\alpha_j)}}$$

Definiendo al vector de parámetros como $\alpha = (\alpha_1, \dots, \alpha_k)$, la representación matricial de la matriz de varianzas y covarianzas es:

$$Var[p] = \frac{1}{A(1+A)} \left(diag(\alpha) - \frac{1}{A} \alpha \alpha' \right)$$

Los momentos de orden superior, se facilitan si se calcula por el uso de los momentos factoriales, esto quiere decir: $E[p_i^{[r]}] = E[p_i(p_i - 1) \dots (p_i - r + 1)]$, de manera conjunta se presentan como:

$$E[\prod_{i=1}^k p_i^{[r_i]}] = \frac{\prod_{i=1}^k \alpha_i^{[r_i]}}{A^{[\sum_{i=1}^k r_i]}}$$

Si bien es un hecho, que cada variable $p_i \sim \mathcal{Be}(\alpha_i; A - \alpha_i)$, pero las combinaciones de sumas de estas también pertenecen a la familia Dirichlet, dicho de otro modo:

Para un $h < k$, si $(p_1, \dots, p_k) \sim \mathcal{Di}(\alpha_1, \dots, \alpha_k)$ entonces:

$$(p_1, \dots, p_h, p_{h+1} + \dots + p_k) \sim \mathcal{Di}(\alpha_1, \dots, \alpha_h, (\alpha_{h+1} + \dots + \alpha_k))$$

Un caso particular de esto es el hecho $p_i \sim \mathcal{Be}(\alpha_i; A - \alpha_i)$, por otro lado las **distribuciones condicionales** en una Dirichlet, tienen una característica particular.

Para un $h < k$, si $(p_1, \dots, p_k) \sim \mathcal{Di}(\alpha_1, \dots, \alpha_k)$ entonces:

$$\frac{1}{\sum_{i=1}^h p_i} (p_1, \dots, p_h) | p_{h+1}, \dots, p_k \sim \mathcal{Di}(\alpha_1, \dots, \alpha_h)$$

Este hecho es importante ya que el vector aleatorio $(1/\sum_{i=1}^h p_i)(p_1, \dots, p_h)$ es independiente de (p_{h+1}, \dots, p_k) . Una transformación muy particular dentro del caso anterior, es la distribución de la proporción de suma, para un $g < h$, si $(p_1, \dots, p_k) \sim \mathcal{Di}(\alpha_1, \dots, \alpha_k)$ entonces:

$$\frac{\sum_{i=1}^g p_i}{\sum_{i=1}^h p_i} \sim \mathcal{Be}(\alpha_1 + \dots + \alpha_g; \alpha_{g+1} + \dots + \alpha_h)$$

La **construcción de una distribución Dirichlet**, se basa puede determinar de dos maneras, la primera con el uso de k variables aleatorias independientes distribuidas mediante una ley Gama $\Lambda_i \sim \mathcal{G}(\alpha_i, \theta_i = 1)$ y la transformación que se aplica es:

$$(p_1, \dots, p_k) = \frac{1}{\sum_{i=1}^k \Lambda_i} (\Lambda_1, \dots, \Lambda_k) \sim \mathcal{Di}(\alpha_1, \dots, \alpha_k)$$

Una segunda manera es mediante el uso de distribuciones Beta, tomando a $k - 1$ variables independientes $q_i \sim \mathcal{Be}(a_i, b_i)$, se procede a un tipo de transformación recursiva:

$$\begin{aligned} p_1 &= q_1 \\ p_i &= q_i \prod_{j=1}^{i-1} (1 - q_j) \quad i = 2, \dots, k - 1 \\ p_k &= \prod_{j=1}^{k-1} (1 - q_j) \end{aligned}$$

Se debe aclarar que $\alpha_i = a_i$ y $b_i = b_{i-1} + \alpha_i$, con $i = 2, \dots, k - 1$. Esta transformación deriva en una $(p_1, \dots, p_k) \sim \mathcal{D}i(\alpha_1, \dots, \alpha_k)$.

A.3.2. Distribución Trapezoidal

La distribución trapezoidal en un inicio fue usada para el análisis de riesgo, pero dado el carácter de enfoque de su aplicación en el campo de los números difusos, solo enfatiza en la distribución trapezoidal estándar, para la cual su *función de densidad* se define como:

$$f(x|\alpha, \beta, \gamma, \delta) = \begin{cases} u \left(\frac{x-\alpha}{\beta-\alpha} \right) & \alpha \leq x < \beta \\ u & \beta \leq x < \gamma \\ u \left(\frac{\delta-x}{\delta-\gamma} \right) & \gamma \leq x < \delta \\ 0 & \text{de otro modo} \end{cases}$$

El término constante “ u ” es igual a $u = 2(\delta + \gamma - \alpha - \beta)^{-1}$, la forma generalizada se desarrollo ampliamente en varios campos de aplicación (Van Dorp & Kotz, 2003), denominada como *distribución trapezoidal generalizada* como depende del vector de parámetros: $\Theta = [\alpha, \beta, \gamma, \delta, \eta_1, \eta_2, \theta]$ y viene dada por:

$$f(x|\Theta) = \begin{cases} K \left(\frac{x-\alpha}{\beta-\alpha} \right)^{\eta_1-1} & \alpha \leq x < \beta \\ K \left\{ (\theta - 1) \frac{\gamma-x}{\gamma-\beta} + 1 \right\} & \beta \leq x < \gamma \\ K \left(\frac{\delta-x}{\delta-\gamma} \right)^{\eta_2-1} & \gamma \leq x < \delta \\ 0 & \text{de otro modo} \end{cases}$$

El espacio parametrico esta dado por: $\alpha < \beta < \gamma < \delta, \eta_1 > 0, \eta_2 > 0, \theta > 0$ La constante K es igual a $K = 2\eta_1\eta_2 / \{2(\beta - \alpha)\eta_2 + (\theta + 1)(\gamma - \beta)\eta_1\eta_2 + 2(\delta - \gamma)\eta_1\}$. El cálculo de la *función de distribución acumulada* se desarrolla como:

1er Caso $t \in (\alpha, \beta)$

$$\begin{aligned} F(x|\Theta) &= \int_{t=\alpha}^{t=x} f(t|\Theta) dt \\ &= \int_{t=\alpha}^{t=x} K \left(\frac{t-\alpha}{\beta-\alpha} \right)^{\eta_1-1} dt \\ &= \left[\frac{K(\beta-\alpha)}{\eta_1} \left(\frac{t-\alpha}{\beta-\alpha} \right)^{\eta_1} \right]_{t=\alpha}^{t=x} \\ &= \frac{K(\beta-\alpha)}{\eta_1} \left(\frac{x-\alpha}{\beta-\alpha} \right)^{\eta_1} - \frac{K(\beta-\alpha)}{\eta_1} \left(\frac{\alpha-\alpha}{\beta-\alpha} \right)^{\eta_1} \\ F(x|\Theta) &= \frac{K(\beta-\alpha)}{\eta_1} \left(\frac{x-\alpha}{\beta-\alpha} \right)^{\eta_1} \end{aligned}$$

2do Caso $t \in (\beta, \gamma)$

$$\begin{aligned}
F(x|\Theta) &= F(\beta|\Theta) + \int_{t=\beta}^{t=x} f(t|\Theta) dt \\
&= F(\beta|\Theta) + \int_{t=\beta}^{t=x} K \left\{ (\theta - 1) \frac{\gamma-x}{\gamma-\beta} + 1 \right\} dt \\
&= F(\beta|\Theta) + K \left[(\theta - 1) \frac{\gamma x - x^2/2}{\gamma-\beta} + x \right]_{t=\beta}^{t=x} \\
&= \frac{K(\beta-\alpha)}{\eta_1} \left(\frac{\beta-\alpha}{\beta-\alpha} \right)^{\eta_1} + K \left[(\theta - 1) \frac{\gamma x - x^2/2}{\gamma-\beta} + x \right] - K \left[(\theta - 1) \frac{\gamma\beta - \beta^2/2}{\gamma-\beta} + \beta \right] \\
&= \frac{K\beta-\alpha}{\eta_1} + K \left[(\theta - 1) \frac{\gamma x - x^2/2}{\gamma-\beta} + x \right] - K \left[(\theta - 1) \frac{\gamma\beta - \beta^2/2}{\gamma-\beta} + \beta \right] \\
&= K \left[\frac{\beta-\alpha}{\eta_1} + (\theta - 1) \frac{\gamma x - x^2/2}{\gamma-\beta} + x - (\theta - 1) \frac{\gamma\beta - \beta^2/2}{\gamma-\beta} - \beta \right] \\
&= K \left[\frac{\beta-\alpha}{\eta_1} + (\theta - 1) \frac{2\gamma x - x^2 - 2\gamma\beta + \beta^2}{2(\gamma-\beta)} + x - \beta \right] \\
&= K \left[\frac{\beta-\alpha}{\eta_1} + (\theta - 1) \frac{2\gamma(x-\beta) + (x+\beta)(x-\beta)}{2(\gamma-\beta)} + x - \beta \right] \\
F(x|\Theta) &= K \left[\frac{\beta-\alpha}{\eta_1} + (x - \beta) \left\{ (\theta - 1) \frac{2\gamma - (x+\beta)}{2(\gamma-\beta)} + 1 \right\} \right]
\end{aligned}$$

3er Caso $t \in (\gamma, \delta)$

$$\begin{aligned}
F(x|\Theta) &= F(\gamma|\Theta) + \int_{t=\gamma}^{t=x} f(t|\Theta) dt \\
&= F(\gamma|\Theta) + \int_{t=\gamma}^{t=x} K \left(\frac{\delta-t}{\delta-\gamma} \right)^{\eta_2-1} dt \\
&= F(\gamma|\Theta) + \left[-\frac{K(\delta-\gamma)}{\eta_2} \left(\frac{\delta-t}{\delta-\gamma} \right)^{\eta_2} \right]_{t=\gamma}^{t=x} \\
&= F(\gamma|\Theta) - \frac{K(\delta-\gamma)}{\eta_2} \left(\frac{\delta-x}{\delta-\gamma} \right)^{\eta_2} + \frac{K(\delta-\gamma)}{\eta_2} \left(\frac{\delta-\gamma}{\delta-\gamma} \right)^{\eta_2} \\
&= K \left[\frac{\beta-\alpha}{\eta_1} + (\gamma - \beta) \left\{ (\theta - 1) \frac{2\gamma - (\gamma+\beta)}{2(\gamma-\beta)} + 1 \right\} \right] - \frac{K(\delta-\gamma)}{\eta_2} \left(\frac{\delta-x}{\delta-\gamma} \right)^{\eta_2} + \frac{K(\delta-\gamma)}{\eta_2} \\
&= K \left[\frac{\beta-\alpha}{\eta_1} + \frac{(\gamma-\beta)(\theta+1)}{2} \right] - \frac{K(\delta-\gamma)}{\eta_2} \left(\frac{\delta-x}{\delta-\gamma} \right)^{\eta_2} + \frac{K(\delta-\gamma)}{\eta_2} \\
F(x|\Theta) &= K \left[\frac{\beta-\alpha}{\eta_1} + \frac{(\gamma-\beta)(\theta+1)}{2} + \frac{\delta-\gamma}{\eta_2} - \frac{\delta-\gamma}{\eta_2} \left(\frac{\delta-x}{\delta-\gamma} \right)^{\eta_2} \right]
\end{aligned}$$

4to Caso $t \in (\delta, +\infty)$

$$\begin{aligned}
F(x|\Theta) &= F(\delta|\Theta) + \int_{t=\delta}^{t=x} f(t|\Theta) dt \\
&= F(\delta|\Theta) + \int_{t=\delta}^{t=x} 0 dt \\
&= F(\delta|\Theta) + 0 \\
&= K \left[\frac{\beta-\alpha}{\eta_1} + \frac{(\gamma-\beta)(\theta+1)}{2} + \frac{\delta-\gamma}{\eta_2} - \frac{\delta-\gamma}{\eta_2} \left(\frac{\delta-\delta}{\delta-\gamma} \right)^{\eta_2} \right] \\
&= K \left[\frac{\beta-\alpha}{\eta_1} + \frac{(\gamma-\beta)(\theta+1)}{2} + \frac{\delta-\gamma}{\eta_2} \right] \\
&= K \left[\frac{2\eta_2(\beta-\alpha) + (\gamma-\beta)(\theta+1)\eta_1\eta_2 + 2(\delta-\gamma)\eta_2}{2\eta_1\eta_2} \right]
\end{aligned}$$

$$F(x|\Theta) = 1$$

Por ende la función de distribución acumulada viene dada por:

$$F(x|\Theta) = \begin{cases} 0 & x < \alpha \\ K \frac{\beta-\alpha}{\eta_1} \left(\frac{x-\alpha}{\beta-\alpha}\right)^{\eta_1} & \alpha \leq x < \beta \\ K \left[\frac{\beta-\alpha}{\eta_1} + (x-\beta) \left\{ (\theta-1) \frac{2\gamma-(x+\beta)}{2(\gamma-\beta)} + 1 \right\} \right] & \beta \leq x < \gamma \\ K \left[\frac{\beta-\alpha}{\eta_1} + \frac{(\gamma-\beta)(\theta+1)}{2} + \frac{\delta-\gamma}{\eta_2} - \frac{\delta-\gamma}{\eta_2} \left(\frac{\delta-x}{\delta-\gamma}\right)^{\eta_2} \right] & \gamma \leq x < \delta \\ 1 & x > \delta \end{cases}$$

Para el cálculo de los momentos ordinarios, se maneja la descomposición de la función original $f(t|\Theta)dt$ en una combinación lineal de otras funciones de densidad de la siguiente manera:

$$f(x|\Theta) = w_1 f_1(x|\Theta_1) + w_2 f_2(x|\Theta_2) + w_3 f_3(x|\Theta_3)$$

Donde los términos w_i son tales que $\sum_i w_i = \lambda\rho + (1-\lambda) + \lambda(1-\rho) = 1$, siendo definidos como:

$$\rho = \frac{(\beta-\alpha)\eta_2\theta}{(\delta-\gamma)\eta_1 + \theta(\beta-\alpha)\eta_2}$$

$$\lambda = \frac{2(\delta-\gamma)\eta_1 + 2\theta(\beta-\alpha)\eta_2}{2(\delta-\gamma)\eta_1 + (\theta+1)(\gamma-\beta)\eta_1\eta_2 + 2\theta(\beta-\alpha)\eta_2}$$

Las funciones de densidad $f_i(x|\Theta_i)$ se expresan de manera excluyente como:

$$f_1(x|\Theta_1) = \frac{\eta_1}{\beta-\alpha} \left(\frac{x-\alpha}{\beta-\alpha}\right)^{\eta_1-1} \quad \alpha \leq x < \beta$$

$$f_2(x|\Theta_2) = \frac{2}{(\theta+1)(\beta-\alpha)^2} \{(1+\theta)x + \theta\gamma - \beta\} \quad \beta \leq x < \gamma$$

$$f_3(x|\Theta_3) = \frac{\eta_2}{\delta-\gamma} \left(\frac{\delta-x}{\delta-\gamma}\right)^{\eta_2-1} \quad \gamma \leq x < \delta$$

Al estar estas funciones definidas en intervalos excluyentes, es posible calcular la esperanza de la variable aleatoria "X" tomando valores esperados a estas nuevas variables desagregadas, es decir:

$$E[X^r|\Theta] = \lambda\rho E[X_1^r|\Theta_1] + (1-\lambda)E[X_2^r|\Theta_2] + \lambda(1-\rho)E[X_3^r|\Theta_3]$$

El cálculo de momentos superiores no es de mucho interés, siendo el énfasis los de 1er orden ($r = 1$) para el caso de defuzzificación, los cuales son:

$$E[X_1|\Theta_1] = \frac{\alpha + \eta_1\beta}{\eta_1 + 1}$$

$$E[X_3|\Theta_3] = \frac{\eta_2\gamma + \delta}{\eta_2 + 1}$$

$$E[X_2|\Theta_2] = \frac{\frac{2}{3}(\theta+1)(\gamma^3-\beta^3)+(\theta\gamma-\beta)(\gamma^2-\beta^2)}{(\gamma-\beta)^2(\theta+1)}$$

Obteniendo finalmente el momento de 1er orden de la distribución trapezoidal generalizada como:

$$E[X|\Theta] = \lambda\rho\left(\frac{\alpha+\eta_1\beta}{\eta_1+1}\right) + (1-\lambda)\left\{\frac{\frac{2}{3}(\theta+1)(\gamma^3-\beta^3)+(\theta\gamma-\beta)(\gamma^2-\beta^2)}{(\gamma-\beta)^2(\theta+1)}\right\} + \lambda(1-\rho)\left(\frac{\eta_2\gamma+\delta}{\eta_2+1}\right)$$

El caso en particular de la distribución trapezoidal estándar, se produce cuando los parámetros $\eta_1 = \eta_2 = 2; \theta = 1$, el cual se adquiere la *esperanza*:

$$E[X|\Theta] = \frac{(\beta-\alpha)(\alpha+2\beta)-3(\beta^2-\gamma^2)+(\delta-\gamma)(2\gamma+\delta)}{3(\delta+\gamma-\beta-\alpha)}$$

De manera análoga se obtiene el momento de 2do orden para el caso de la distribución trapezoidal estándar:

$$E[X^2|\Theta] = \frac{\beta-\alpha}{\delta+\gamma-\beta-\alpha}\left(\frac{(\alpha+\beta)^2}{6}+\frac{\beta^2}{3}\right) + \frac{1}{\delta+\gamma-\beta-\alpha}\left(\frac{2}{3}(\gamma^2-\beta^2)\right) + \frac{\delta-\gamma}{\delta+\gamma-\beta-\alpha}\left(\frac{\gamma^2}{3}+\frac{(\gamma+\delta)^2}{6}\right)$$

A.4. Metodología BOX & JENKINS: Modelo ARIMA

En 1970, Box & Jenkins desarrollaron un marco metodológico destinado a identificar, estimar y diagnosticar modelos dinámicos de series temporales en los que la variable tiempo juega un papel fundamental.

Para la descripción breve de la metodología de Box & Jenkins, se hace uso de procesos lineales, comenzando por el proceso de ruido blanco Z_t . Definido de esta manera Z_t , en base a este, se denominan como procesos autoregresivos $AR(p)$ expresándolos de la siguiente forma:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t$$

Donde ε_t es un proceso de ruido blanco con media cero, y varianza constante σ_ε^2 . Se define también los modelos de medias móviles de orden q , o proceso $MA(q)$ que viene dado por:

$$Z_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

A la composición de procesos autoregresivos y un proceso de medias móviles se lo denomina $ARMA(p, q)$ donde p indica el retardo máximo de la parte autoregresiva, y q señala el correspondiente a la parte de las medias móviles, este ultimo esta formulado como:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

Es muy común el uso de operadores de tipo retraso, así que estos se manejan como Polinomios de rezago, denotados para este caso como: $\Phi(B)$ y $\Theta(B)$, ahora la ecuación anterior se puede reescribir mediante:

$$\Phi(B)Z_t = \Theta(B)\varepsilon_t$$

Además de los modelos como tal, es necesario enfatizar en el operador diferencia, denotados como " ∇ ", se tienen de diferente orden, de 1er orden: $\nabla Z_t = Z_t - Z_{t-1}$, de segundo orden $\nabla^2 Z_t = \nabla Z_t - \nabla Z_{t-1}$, y de manera análoga se define para un orden superior " d ", como: $\nabla^d Z_t = \nabla^{d-1} Z_t - \nabla^{d-1} Z_{t-1}$. Tras estos conceptos vertidos, un modelo $ARIMA(p, d, q)$ como tal está definido mediante:

$$\Phi(B)\nabla^d Z_t = \Theta(B)\varepsilon_t$$

A.5. DO-FILE STATA 12.0

Se desarrollo una serie de rutinas, para facilitar el cálculo y procesamiento de los datos, todas las rutinas están elaboradas para el paquete estadístico STATA 12.0, habiendo algunas variantes para versiones inferiores.

Los programas a continuación mencionados, desarrollan la siguiente secuencia de pasos, 1ro la creación de variables y relaciones, 2do la estimación y reajuste de la matriz de Markov, 3ro la construcción de los números difusos, 4to la estimación del modelo ARIMA (que tiene por defecto el software), y por último la comparación entre los modelos de Markov y ARIMA.

```
*****.
**dofile stata 12.0
**modelo de Markov
*clear
*matrix drop _all
*program drop _all
*ereturn clear
**creacion de funcion ECM
*program e_c_m
*quietly {
*capture drop ee`2'
*gen ee`2'=(`1'-`2')^2
*total ee`2'
*matrix ee1_s=e( b)
*matrix ee1_n=e(_N)
*}
*disp as text " ECM=" as result ee1_s[1,1]/ee1_n[1,1]
*capture drop ee`2'
*matrix drop ee1_s ee1_n
*end
**Importación de datos.
*import excel "C:\Users\RONAL\Desktop\Docs_tesis\series y datos\"
* "Nueva carpeta\datos_afps_potosi.xlsx", sheet("Hoja1") firstrow
*keep y_t
**creación de variables*
*local n_dim = 14
*label variable y_t "Total recaudado [Bs]"
*egen int t_2 = fill(449(1)450)
*tsset t_2, monthly
*label variable t_2 "Periodo de recaudación"
**transformaciones
*gen logyt = ln(y_t)
*gen Slogyt = d1.logyt
*label variable Slogyt "W_t"
*ai estado Y(t)
*aj estado Y(t-1)
*xtile ai = Slogyt, nq(`n_dim')
*gen aj = L1.ai
**matriz de dos estados
*quiet tabulate aj ai, matcell(Hkv)
**priori
*matrix input kir01=(1.0 0.5 0 0 0 0 0 0 0 0 0 0 0 0 )
*matrix input kir02=(0.5 1.0 0.5 0 0 0 0 0 0 0 0 0 0 0 )
*matrix input kir03=(0 0.5 1.0 0.5 0 0 0 0 0 0 0 0 0 0 )
*matrix input kir04=(0 0 0.5 1.0 0.5 0 0 0 0 0 0 0 0 0 )
*matrix input kir05=(0 0 0 0.5 1.0 0.5 0 0 0 0 0 0 0 0 )
*matrix input kir06=(0 0 0 0 0.5 1.0 0.5 0 0 0 0 0 0 0 )
*matrix input kir07=(0 0 0 0 0 0.5 1.0 0.5 0 0 0 0 0 0 )
*matrix input kir08=(0 0 0 0 0 0 0.5 1.0 0.5 0 0 0 0 0 )
*matrix input kir09=(0 0 0 0 0 0 0 0.5 1.0 0.5 0 0 0 0 )
```

```

*matrix input kir10=(0 0 0 0 0 0 0 0 0.5 1.0 0.5 0 0 0 )
*matrix input kir11=(0 0 0 0 0 0 0 0 0 0.5 1.0 0.5 0 0 )
*matrix input kir12=(0 0 0 0 0 0 0 0 0 0 0.5 1.0 0.5 0 )
*matrix input kir13=(0 0 0 0 0 0 0 0 0 0 0 0.5 1.0 0.5)
*matrix input kir14=(0 0 0 0 0 0 0 0 0 0 0 0 0.5 1.0)
*matrix kir =(kir01\kir02\kir03\kir04\kir05\kir06
*
// \kir07\kir08\kir09\kir10\kir11\kir12\kir13\kir14)
**estimación de markov
*matrix Hkv=Hkv+kir
*matrix Mkv = (Hkv*inv( diag( Hkv* J(`n_dim',1,1) ) ))'
**matriz de estados simples
*quiet tabulate aj , matcell(Hk1)
*matrix input kirp =(0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5)
**estimación de Multinomial
*matrix Hk1=Hk1
*matrix Mk1 = (Hk1*inv( diag( Hk1* J(`n_dim',1,1) ) ))'
**listado de matrices
*matrix list Mkv, format(%12.2f)
*matrix list Mk1, format(%12.2f)
**matriz de centroides
*tabstat Slogyt,statistics(min max p25 p50 p75 mean sd)
*
// by(ai) columns(statistics) save
*matrix M_qnt0=r(Stat1)
*forvalues i = 2(1)`n_dim' {
*matrix M_qnt0=(M_qnt0,r(Stat`i'))
*}
*matrix M_qnt=M_qnt0'
**listado de matrices
*matrix list M_qnt, format(%12.2f)
*numeros difusos.
*matrix Num_1=J(`n_dim',2,0)
*matrix Num_2=J(`n_dim',4,0)
*forvalues i = 1(1)`n_dim' {
*matrix Num_2[`i',1]= (M_qnt[`i',6]-3*M_qnt[`i',7])
*matrix Num_2[`i',2]= (M_qnt[`i',6]-1*M_qnt[`i',7])
*matrix Num_2[`i',3]= (M_qnt[`i',6]+1*M_qnt[`i',7])
*matrix Num_2[`i',4]= (M_qnt[`i',6]+3*M_qnt[`i',7])
*}
**listado de matrices
*matrix list Num_2, format(%12.2f)
**matriz ajustada
*matrix Bconf=J(`n_dim',1,0)
*forvalues i = 1(1)`n_dim' {
*matrix set01=diag(Mkv[`i'..'i', 1..`n_dim']) *Mk1
*matrix Bconf=(Bconf,set01/trace(diag(set01)) )
*}
*matrix BCONF=(Bconf[1..`n_dim',2..(`n_dim'+1)])'
**listado de matrices
*matrix list BCONF, format(%12.2f)
**valores esperados
*gen L_logyt = L1.logyt
*mkmat L_logyt aj, matrix(Y_obs)
*matrix E_num = BCONF*Num_2
**listado de matrices
*matrix list E_num, format(%12.2f)
**asignación de valores esperados
*matrix Y_fact=J(_N,`n_dim',0)
*forvalues i = 1(1)175 {
*forvalues j = 1(1)`n_dim' {
*matrix Y_fact[`i',`j'] = Y_fact[`i',`j'] + (Y_obs[`i',2]==`j')
*}
*}
*matrix E_num2=Y_fact*E_num
**datos fuzificados
*matrix Y_obs2=J(_N,4,0)
*forvalues i = 1(1)175 {
*forvalues j = 1(1)4 {
*matrix Y_obs2[`i',`j'] = Y_obs[`i',1]
*}
*}

```

```

*}
**matriz final
*matrix Y_fuzzy=Y_obs2+E_num2
*svmat Y_fuzzy
*
**defuzzificadores
**max heigth
*gen y_fuzzy=(Y_fuzzy2+Y_fuzzy3)/2
**average
*gen y_mean =((Y_fuzzy2-Y_fuzzy1)*(2*Y_fuzzy2+Y_fuzzy1)-3*
// (Y_fuzzy2^2-Y_fuzzy3^2)+(Y_fuzzy4-Y_fuzzy3)*
// (Y_fuzzy4+2*Y_fuzzy3))/(3*(Y_fuzzy4+Y_fuzzy3-Y_fuzzy2-Y_fuzzy1))
**re transformando
*gen y1_t=exp(y_ar)
*gen y2_t=exp(y_mean)
*gen y3_t=exp(y_fuzzy)
**errores
*e_c_m logyt y_ar
*e_c_m logyt y_mean
*e_c_m logyt y_fuzzy
*****.

```

A.6. Datos y Resultados Obtenidos

Los resultados obtenidos, se dividieron en dos partes, la primera centrada a la exposición de las matrices y números difusos obtenidos. La segunda parte expone de manera modesta los resultados obtenidos para el modelo ARIMA, y las pruebas pertinentes al caso de aplicación.

TABLA 8. SERIE DE DATOS MANEJADOS

y_t	t	y_t	t	y_t	t	y_t	t
1253,9867	1997m6	2048,1508	2001m8	3313,0763	2005m10	5669,8609	2009m12
1500,0089	1997m7	1973,7917	2001m9	3643,6943	2005m11	6474,3091	2010m1
1612,0723	1997m8	2220,0012	2001m10	3374,9878	2005m12	5118,0809	2010m2
1510,1815	1997m9	1858,7059	2001m11	3531,3723	2006m1	5193,6519	2010m3
1737,1741	1997m10	1783,6557	2001m12	3222,8088	2006m2	6166,0547	2010m4
1604,0013	1997m11	1894,6009	2002m1	3761,4336	2006m3	6026,727	2010m5
1623,5373	1997m12	1676,2206	2002m2	3596,6557	2006m4	5626,479	2010m6
1494,7127	1998m1	1947,1854	2002m3	4218,106	2006m5	4624,405	2010m7
1601,9382	1998m2	2218,8527	2002m4	3468,4727	2006m6	6454,3748	2010m8
1565,7448	1998m3	1849,909	2002m5	4789,1237	2006m7	6047,6784	2010m9
1505,2427	1998m4	1839,021	2002m6	4080,9593	2006m8	8269,5768	2010m10
1341,2911	1998m5	2074,9517	2002m7	4275,4983	2006m9	7018,1347	2010m11
1951,7624	1998m6	2953,6191	2002m8	4398,7427	2006m10	6593,146	2010m12
2009,9302	1998m7	1896,1253	2002m9	3829,5631	2006m11	5549,2482	2011m1
1709,6408	1998m8	2018,9929	2002m10	5320,1135	2006m12	7105,6548	2011m2
1901,1588	1998m9	2118,0909	2002m11	4511,8918	2007m1	7954,8183	2011m3
1972,9462	1998m10	2228,4138	2002m12	4549,3309	2007m2	7614,443	2011m4
1741,3343	1998m11	2209,762	2003m1	4693,5206	2007m3	8170,1685	2011m5
1841,1602	1998m12	3049,9749	2003m2	4468,9407	2007m4	11151,643	2011m6
1827,8819	1999m1	2250,0347	2003m3	4661,2195	2007m5	8799,2836	2011m7
1444,5059	1999m2	2042,9952	2003m4	4059,8251	2007m6	9179,3959	2011m8
1982,042	1999m3	2345,0717	2003m5	3984,2166	2007m7	9110,9714	2011m9
1812,5049	1999m4	2621,2848	2003m6	4586,0805	2007m8	10077,715	2011m10
1868,5759	1999m5	2669,9859	2003m7	4880,6767	2007m9	9719,6947	2011m11
1670,2421	1999m6	2648,785	2003m8	6774,4912	2007m10	10883,15	2011m12
1788,7421	1999m7	2291,9594	2003m9	5022,9606	2007m11		
1652,1212	1999m8	2926,9298	2003m10	5077,9932	2007m12		
2058,2193	1999m9	2498,0977	2003m11	4798,4091	2008m1		
1637,9708	1999m10	2634,4203	2003m12	4846,097	2008m2		
1800,9369	1999m11	3020,967	2004m1	5035,7716	2008m3		
1588,6865	1999m12	2233,6883	2004m2	5103,2644	2008m4		
1643,0522	2000m1	2352,2753	2004m3	5482,8053	2008m5		
1366,2954	2000m2	2362,8994	2004m4	5532,7773	2008m6		
1777,3446	2000m3	2355,6254	2004m5	6094,5556	2008m7		
1723,0687	2000m4	2627,6508	2004m6	6316,7633	2008m8		
2055,9484	2000m5	3054,7432	2004m7	5813,5821	2008m9		
1710,0093	2000m6	2807,5933	2004m8	6882,2452	2008m10		
2228,5159	2000m7	3030,7763	2004m9	5624,3275	2008m11		
2230,5711	2000m8	2676,8774	2004m10	6235,7933	2008m12		
1796,8256	2000m9	2592,0264	2004m11	6216,2759	2009m1		
1984,5213	2000m10	3092,2263	2004m12	4312,9037	2009m2		
1863,1259	2000m11	2453,9741	2005m1	5616,573	2009m3		
2103,2096	2000m12	2337,98	2005m2	5066,0109	2009m4		
1817,5822	2001m1	2670,0711	2005m3	4539,4856	2009m5		
1791,3623	2001m2	2771,4307	2005m4	5073,4904	2009m6		
2119,2082	2001m3	2760,1022	2005m5	6692,4054	2009m7		
2483,3565	2001m4	3169,9123	2005m6	6059,497	2009m8		
2053,7783	2001m5	2769,7014	2005m7	6138,9285	2009m9		
2199,2568	2001m6	3062,2365	2005m8	5093,5999	2009m10		
1919,9437	2001m7	2848,148	2005m9	7417,3804	2009m11		

Antes de presentar los datos es necesario listar los datos de manera exhaustiva, los mismos fueron recabados de la El origen serie analizada en su integridad se encuentra disponible en el portal de UDAPE (<http://www.udape.gob.bo/>), estos datos se ven en la tabla anterior.

A.6.1. Para el Modelo de Markov

De manera secuencial se describen el contenido de cada una de las matrices obtenidas durante todo el proceso de estimación del Modelo de Markov Difuso. Se debe considerar que solo se expondrán resultados, no así las sentencias, ya que las mismas se presentaron en el Anexo A.5.

```
*MATRIZ DE MARKOV INICIAL
Mkv[14,14]
  c1      c2      c3      c4      c5      c6      c7      c8      c9      c10     c11     c12     c13     c14
r1 0.000 0.000 0.000 0.077 0.077 0.000 0.077 0.077 0.231 0.154 0.077 0.077 0.077 0.077
r2 0.000 0.000 0.000 0.000 0.167 0.000 0.167 0.000 0.000 0.083 0.083 0.000 0.250 0.250
r3 0.000 0.000 0.000 0.000 0.000 0.231 0.000 0.077 0.231 0.154 0.000 0.077 0.077 0.154
r4 0.000 0.000 0.000 0.083 0.000 0.000 0.083 0.167 0.000 0.250 0.083 0.167 0.167 0.000
r5 0.000 0.154 0.000 0.000 0.000 0.000 0.077 0.000 0.231 0.077 0.077 0.308 0.000 0.077
r6 0.000 0.000 0.167 0.000 0.083 0.083 0.000 0.000 0.000 0.000 0.250 0.083 0.250 0.083
r7 0.250 0.000 0.000 0.083 0.083 0.000 0.083 0.167 0.000 0.083 0.167 0.083 0.000 0.000
r8 0.000 0.231 0.154 0.154 0.077 0.077 0.077 0.077 0.000 0.077 0.000 0.000 0.077 0.000
r9 0.000 0.000 0.167 0.083 0.000 0.167 0.167 0.083 0.167 0.000 0.000 0.083 0.000 0.083
r10 0.000 0.000 0.231 0.077 0.308 0.154 0.077 0.077 0.000 0.000 0.000 0.000 0.000 0.077
r11 0.000 0.182 0.091 0.000 0.091 0.091 0.091 0.182 0.000 0.000 0.000 0.091 0.000 0.182
r12 0.154 0.154 0.077 0.154 0.077 0.000 0.000 0.077 0.077 0.077 0.154 0.000 0.000 0.000
r13 0.250 0.083 0.083 0.083 0.000 0.167 0.083 0.000 0.000 0.083 0.083 0.083 0.000 0.000
r14 0.417 0.167 0.083 0.167 0.083 0.000 0.000 0.083 0.000 0.000 0.000 0.000 0.000 0.000

*MATRIZ ESTIMADA MULTINOMIAL
Mk1[14,1]
  c1
r1 0.075
r2 0.069
r3 0.075
r4 0.069
r5 0.075
r6 0.069
r7 0.069
r8 0.075
r9 0.069
r10 0.075
r11 0.064
r12 0.075
r13 0.069
r14 0.069

*MATRIZ DE ESTADISTICOS BASICOS
M_qnt[14,7]
  min    max    p25    p50    p75    mean    sd
Slogyt -0.44 -0.20 -0.30 -0.24 -0.23 -0.27 0.07
Slogyt -0.20 -0.16 -0.19 -0.18 -0.17 -0.18 0.01
Slogyt -0.16 -0.12 -0.14 -0.14 -0.12 -0.14 0.01
Slogyt -0.11 -0.08 -0.10 -0.09 -0.08 -0.09 0.01
Slogyt -0.08 -0.04 -0.07 -0.06 -0.05 -0.06 0.01
Slogyt -0.04 -0.01 -0.03 -0.02 -0.01 -0.02 0.01
Slogyt -0.01 0.01 -0.00 0.00 0.01 0.00 0.01
Slogyt 0.01 0.04 0.02 0.03 0.04 0.03 0.01
Slogyt 0.04 0.06 0.05 0.05 0.06 0.05 0.01
Slogyt 0.06 0.10 0.07 0.07 0.10 0.08 0.01
Slogyt 0.10 0.13 0.11 0.11 0.12 0.11 0.01
Slogyt 0.13 0.16 0.14 0.14 0.15 0.14 0.01
Slogyt 0.17 0.26 0.17 0.20 0.26 0.21 0.04
Slogyt 0.28 0.38 0.31 0.33 0.34 0.33 0.03

*NUMEROS DIFUSOS APROXIMADOS
Num_2[14,4]
  c1      c2      c3      c4
r1 -0.480 -0.343 -0.206 -0.069
r2 -0.216 -0.192 -0.168 -0.144
r3 -0.177 -0.150 -0.122 -0.095
```

```

r4 -0.127 -0.104 -0.081 -0.058
r5 -0.093 -0.070 -0.047 -0.023
r6 -0.059 -0.035 -0.011 0.012
r7 -0.019 -0.005 0.010 0.024
r8 -0.001 0.018 0.037 0.056
r9 0.030 0.044 0.059 0.074
r10 0.039 0.067 0.095 0.123
r11 0.088 0.105 0.122 0.138
r12 0.116 0.135 0.154 0.173
r13 0.086 0.170 0.254 0.338
r14 0.247 0.302 0.357 0.413
*MATRIZ DE MARKOV AJUSTADA
BCONF[14,14]
      r1      r2      r3      r4      r5      r6      r7      r8      r9      r10     r11     r12     r13     r14
c1 0.00 0.00 0.00 0.08 0.08 0.00 0.08 0.08 0.23 0.16 0.07 0.08 0.08 0.08
c1 0.00 0.00 0.00 0.00 0.18 0.00 0.16 0.00 0.00 0.09 0.08 0.00 0.25 0.25
c1 0.00 0.00 0.00 0.00 0.00 0.23 0.00 0.08 0.23 0.16 0.00 0.08 0.08 0.15
c1 0.00 0.00 0.00 0.08 0.00 0.00 0.08 0.17 0.00 0.26 0.07 0.17 0.16 0.00
c1 0.00 0.15 0.00 0.00 0.00 0.00 0.08 0.00 0.23 0.08 0.07 0.33 0.00 0.08
c1 0.00 0.00 0.18 0.00 0.09 0.08 0.00 0.00 0.00 0.23 0.09 0.25 0.08
c1 0.26 0.00 0.00 0.08 0.09 0.00 0.08 0.17 0.00 0.09 0.15 0.09 0.00 0.00
c1 0.00 0.22 0.16 0.15 0.08 0.07 0.07 0.08 0.00 0.08 0.00 0.07 0.00
c1 0.00 0.00 0.18 0.08 0.00 0.16 0.16 0.09 0.16 0.00 0.00 0.09 0.00 0.08
c1 0.00 0.00 0.24 0.07 0.32 0.15 0.07 0.08 0.00 0.00 0.00 0.00 0.07
c1 0.00 0.18 0.09 0.00 0.09 0.09 0.09 0.19 0.00 0.00 0.00 0.09 0.18
c1 0.16 0.15 0.08 0.15 0.08 0.00 0.00 0.08 0.08 0.08 0.14 0.00 0.00
c1 0.26 0.08 0.09 0.08 0.00 0.16 0.08 0.00 0.00 0.09 0.07 0.09 0.00
c1 0.43 0.16 0.09 0.16 0.09 0.00 0.00 0.09 0.00 0.00 0.00 0.00 0.00
*VALORES ESPERADOS DIFUSOS
E_num[14,4]
      c1      c2      c3      c4
c1 0.03 0.06 0.09 0.12
c1 0.07 0.12 0.16 0.21
c1 0.05 0.08 0.11 0.15
c1 0.04 0.07 0.10 0.13
c1 0.04 0.06 0.08 0.10
c1 0.03 0.07 0.11 0.15
c1 -0.12 -0.07 -0.02 0.03
c1 -0.10 -0.07 -0.04 -0.02
c1 -0.02 0.00 0.03 0.05
c1 -0.07 -0.05 -0.02 0.00
c1 -0.02 0.01 0.04 0.07
c1 -0.13 -0.09 -0.05 -0.01
c1 -0.16 -0.11 -0.05 -0.00
c1 -0.28 -0.21 -0.14 -0.07

```

A.6.2. Para el Modelo Arima

A continuación se expone el proceso que se desarrollo para la *identificación del Modelo ARIMA(p, d, q) SARIMA(P, D, Q)*, en primera instancia, dado que es una serie mensual, se procede ha de aplicar una diferencia estacional $D = 1$, con $s = 12$ y posteriormente una diferencia ordinaria $d = 1$, verificándose que la serie es estacionaria:

```

Augmented Dickey-Fuller test for unit root          Number of obs   =       160
-----
                Test              ----- Interpolated Dickey-Fuller -----
                Statistic          1% Critical   5% Critical   10% Critical
                Value              Value         Value         Value
-----
Z(t)            -14.288            -3.490        -2.886        -2.576
-----
MacKinnon approximate p-value for Z(t) = 0.0000
-----
D2S12.logyt |      Coef.  Std. Err.   t    P>|t|    [95% Conf. Interval]
-----+-----
      logyt |
LDS12. | -1.899271   .1329249  -14.29  0.000   -2.161822   -1.636719
LD2S12. |  .2639355   .0766976   3.44   0.001   .1124432    .4154278
      _cons |  .0024927   .0153701   0.16   0.871   -.0278662    .0328516
-----+-----

```

De este modo se identifico el orden de diferenciación $d = 1$ y $D = 1$, como ya se tiene una serie estacionaria, se procede a al identificación del los parámetros (p, q) y (P, Q) tanto para la parte ordinaria como estacional respectivamente. Esto se hace con las funciones de Autocorrelacion y Correlacion Parcial:

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	-0.4958	-0.4976	40.565	0.0000	---		---	---		---
2	0.0541	-0.2639	41.051	0.0000	---		---	---		---
3	-0.1095	-0.2942	43.053	0.0000	---		---	---		---
4	0.1100	-0.1438	45.086	0.0000	---		---	---		---
5	-0.0897	-0.1829	46.448	0.0000	---		---	---		---
6	0.1513	0.0335	50.348	0.0000	---		---	---		---
7	-0.1321	-0.0281	53.338	0.0000	---		---	---		---
8	0.0352	-0.0596	53.551	0.0000	---		---	---		---
9	-0.0036	-0.0142	53.553	0.0000	---		---	---		---
10	-0.0081	-0.0657	53.565	0.0000	---		---	---		---
11	0.2847	0.4816	67.824	0.0000	---		---	---		---
12	-0.4901	-0.2444	110.36	0.0000	---		---	---		---
13	0.2013	-0.2126	117.59	0.0000	---		---	---		---
14	0.0182	-0.0854	117.65	0.0000	---		---	---		---
15	0.0306	-0.1291	117.82	0.0000	---		---	---		---
16	-0.0662	-0.0943	118.62	0.0000	---		---	---		---
17	0.1484	0.0270	122.65	0.0000	---		---	---		---
18	-0.2218	-0.0596	131.73	0.0000	---		---	---		---
19	0.1346	-0.0523	135.1	0.0000	---		---	---		---
20	-0.0623	-0.0715	135.82	0.0000	---		---	---		---
21	-0.0004	-0.2226	135.82	0.0000	---		---	---		---
22	0.0837	-0.0059	137.15	0.0000	---		---	---		---
23	-0.0948	0.3349	138.87	0.0000	---		---	---		---
24	0.0313	-0.1492	139.06	0.0000	---		---	---		---
25	0.0729	0.0185	140.09	0.0000	---		---	---		---
26	-0.1455	-0.1265	144.22	0.0000	---		---	---		---
27	0.0996	-0.0792	146.18	0.0000	---		---	---		---
28	-0.0013	-0.1424	146.18	0.0000	---		---	---		---
29	-0.1004	-0.1465	148.19	0.0000	---		---	---		---
30	0.1233	0.0070	151.25	0.0000	---		---	---		---
31	-0.0128	0.0324	151.29	0.0000	---		---	---		---
32	-0.0372	-0.0516	151.57	0.0000	---		---	---		---
33	0.0566	-0.1376	152.23	0.0000	---		---	---		---
34	-0.1294	-0.1082	155.7	0.0000	---		---	---		---
35	0.1231	0.3345	158.87	0.0000	---		---	---		---
36	-0.0028	0.1230	158.88	0.0000	---		---	---		---
37	-0.0781	0.0538	160.17	0.0000	---		---	---		---
38	0.0694	-0.0991	161.2	0.0000	---		---	---		---
39	-0.0207	-0.1605	161.29	0.0000	---		---	---		---
40	-0.0318	-0.0788	161.52	0.0000	---		---	---		---

Al observar los valores de las funciones de Autocorrelación y la Auto correlación parcial se puede identificar cuatro modelos factibles:

- Modelo I: $ARIMA(0,1,1) SARIMA(1,1,1)$
- Modelo II: $ARIMA(0,1,1) SARIMA(1,1,0)$
- Modelo III: $ARIMA(0,1,1) SARIMA(0,1,1)$
- Modelo IV: $ARIMA(0,1,1) SARIMA(0,1,0)$

A continuación se procede a la estimación de cada uno de los modelos por separado, para luego seleccionar al mas idóneo.

arima logyt, arima(0,1,1) sarima(1,1,1,12)
 Sample: 1998m7 - 2011m12
 Log likelihood = 96.40765
 Number of obs = 162
 Wald chi2(3) = 173.96
 Prob > chi2 = 0.0000

DS12.logyt		Coef.	OPG Std. Err.	z	P> z	[95% Conf. Interval]	
logyt	_cons	.0008803	.0005023	1.75	0.080	-.0001042	.0018649
ARMA	ma						
	L1.	-.7459904	.0569043	-13.11	0.000	-.8575208	-.6344601
ARMA12	ar						
	L1.	-.1504475	.1096534	-1.37	0.170	-.3653641	.0644691
	ma						
	L1.	-1.000087	162.2888	-0.01	0.995	-319.0803	317.0801
/sigma		.1192679	9.680398	0.01	0.495	0	19.0925

arima logyt, arima(0,1,1) sarima(1,1,0,12)
 Sample: 1998m7 - 2011m12
 Log likelihood = 70.02071
 Number of obs = 162
 Wald chi2(2) = 307.78
 Prob > chi2 = 0.0000

DS12.logyt		Coef.	OPG Std. Err.	z	P> z	[95% Conf. Interval]	
logyt	_cons	.0010886	.0021034	0.52	0.605	-.003034	.0052112
ARMA	ma						
	L1.	-.7457278	.0527972	-14.12	0.000	-.8492084	-.6422473
ARMA12	ar						
	L1.	-.541178	.0690691	-7.84	0.000	-.6765509	-.4058051
/sigma		.1546478	.0077483	19.96	0.000	.1394613	.1698342

estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	162	.	70.02071	4	-132.0414	-119.691

arima logyt, arima(0,1,1) sarima(0,1,1,12)

Sample: 1998m7 - 2011m12
 Log likelihood = 94.83676
 Number of obs = 162
 Wald chi2(2) = 170.43
 Prob > chi2 = 0.0000

DS12.logyt		Coef.	OPG Std. Err.	z	P> z	[95% Conf. Interval]	
logyt	_cons	.0008837	.0005724	1.54	0.123	-.0002382	.0020057
ARMA	ma						
	L1.	-.74692	.0573569	-13.02	0.000	-.8593374	-.6345027
ARMA12	ma						
	L1.	-1.00001	1443.033	-0.00	0.999	-2829.294	2827.294
/sigma		.1217139	87.82052	0.00	0.499	0	172.2468

arima logyt, arima(0,1,1) sarima(0,1,0,12)
 Sample: 1998m7 - 2011m12
 Log likelihood = 45.63315
 Number of obs = 162
 Wald chi2(1) = 170.29
 Prob > chi2 = 0.0000

DS12.logyt		Coef.	OPG Std. Err.	z	P> z	[95% Conf. Interval]	
logyt	_cons	.0009924	.0036408	0.27	0.785	-.0061435	.0081283
ARMA	ma						
	L1.	-.7577296	.0580659	-13.05	0.000	-.8715366	-.6439225

/sigma	.182089	.0094567	19.26	0.000	.1635543	.2006238

estat ic						

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	162	.	45.63315	3	-85.26631	-76.00352

Examinados la significancia de los coeficientes estimados por un lado, y bajo el uso del criterio Akaike, se elige el modelo II: *ARIMA(0,1,1) SARIMA(1,1,0)*, identificado este modelo procedemos a la validación del modelo, verificando que los errores sean un ruido blanco y los de normalidad.

predict errores, resid
corrgram errores

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	-0.0156	-0.0157	.04009	0.8413						
2	0.0544	0.0549	.5315	0.7666						
3	-0.0868	-0.0863	1.7912	0.6169						
4	0.1206	0.1185	4.2357	0.3750						
5	-0.0105	0.0012	4.2542	0.5134						
6	0.1642	0.1509	8.8442	0.1825	-			-		
7	-0.0894	-0.0733	10.215	0.1767						
8	0.0377	0.0168	10.461	0.2342						
9	-0.0382	-0.0093	10.714	0.2958						
10	0.1164	0.0821	13.083	0.2191						
11	0.0705	0.1214	13.957	0.2354						
12	-0.1289	-0.1933	16.9	0.1534	-			-		
13	-0.0273	0.0104	17.033	0.1978						
14	-0.0170	-0.0307	17.085	0.2517						
15	0.0236	-0.0009	17.185	0.3079						
16	-0.0206	-0.0346	17.263	0.3688						
17	0.0479	0.0507	17.684	0.4090						
18	-0.1550	-0.1362	22.116	0.2269	-			-		
19	0.0304	-0.0009	22.288	0.2702						
20	-0.0705	-0.0577	23.217	0.2783						
21	0.0134	-0.0600	23.251	0.3307						
22	0.0656	0.1831	24.067	0.3437				-		
23	0.0578	0.1299	24.706	0.3656				-		
24	-0.2561	-0.3448	37.333	0.0405	-			-		
25	0.0021	-0.0415	37.334	0.0537						
26	-0.1465	-0.1662	41.526	0.0275	-			-		
27	0.0797	0.0402	42.776	0.0276						
28	-0.0218	0.0004	42.87	0.0358						
29	0.0084	0.0457	42.885	0.0467						
30	0.0432	0.1699	43.26	0.0556						
31	0.0338	0.0249	43.492	0.0675				-		
32	-0.0575	-0.1183	44.169	0.0745						
33	0.0536	-0.0572	44.761	0.0831						
34	-0.0745	0.0386	45.913	0.0834						
35	0.0352	0.1977	46.172	0.0980				-		
36	-0.0227	-0.1350	46.281	0.1172				-		
37	0.0022	-0.0310	46.282	0.1409						
38	0.0156	-0.0645	46.334	0.1662						
39	0.0061	-0.0494	46.342	0.1953						
40	-0.0158	-0.0045	46.396	0.2255						

wntestq errores

Portmanteau test for white noise

Portmanteau (Q) statistic = 46.3959
Prob > chi2(40) = 0.2255

swilk errores

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
errores	162	0.99408	0.736	-0.698	0.75740

BIBLIOGRAFÍA

- Anderson, T., & Goodman, L. (1955). Statistical Inference about Markov Chains. *the Annals of Mathematical Statistics* , 89-110.
- Aponte, G., Ferrufino, R., & Jemio, L. C. (2008). *Pensiones y Jubilacion en Bolivia*. La Paz, Bolivia: HOLDING.
- Arutchelvan, G., Sivasta, S., & Jaganathan, R. (2010). Inaccuracy Minimization by Partitioning Fuzzy Data Sets – Validation of an Analytical Methodology. *International Journal of Computer Science and Information Security*, vol 8 , 275-280.
- Atkinson, A. B. (1970). On the Measurement of Inequality. *Journal of Economic Theory*, vol 2 , 244-263.
- Box, G., & Tiao, G. (1992). *Bayesian Inference in Statistical Analysis*. Canada: John Wiley & Sons, Inc.
- Chen, S. (1996). Forecasting enrollments based on fuzzy time series. *Fuzzy Sets and Systems*, vol 81 , 311-319.
- Dubois, D., & Prade, H. (1979). Fuzzy real algebra: Some results. *Fuzzy Sets and Systems* , 327-348.
- Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, 3th Edition*, vol. I. New York: John Wiley & Sons, Inc.
- Grimmett, G., & Stirzaker, D. (1993). *Probability and Random Processes: Problems and Solutions*. Oxford: Oxford University Press.
- Jeffreys, H. (1961). *Theory of Probability, 3rd ed.* Oxford: Oxford Classic Texts in the Physical Sciences. Oxford Univ. Press,.
- Jilani, T., Burney, A. S., & Ardil, C. (2008). A New Quantile Based Fuzzy Time Series: Forecasting Model. *World Academy of Science, Engineering and Technology*, vol 42 , 99-105.
- Klir, G., & Yuan, B. (1995). *Fuzzy sets and Fuzzy Logic: Theory and Applications*. New Jersey: Prentice Hall Inc.
- Kwakernaak, H. (1978). Fuzzy random variables: Definition and theorems. *Informatic Science vol 15* , 1-29.
- Lindley, D. (1964). The Bayesian analysis of contingency tables. *The Annals of Mathematical*, vol 35 , 1622-1643.
- Liu, X., & Pedrycz, W. (2009). *Axiomatic Fuzzy Set Theory and Its Applications*. Dalian: Springer-Verlag Berlin Heidelberg.

- Marquiegui, R. (1994). *Procesos Estocasticos, vol I,II*. La Paz, Bolivia: Universidad Mayor de San Andres, Carrera de Estadistica.
- Morillas, A. (2002). *Introduccion al Analisis de Datos Difusos*. Colombia: Universidad de Malaga.
- Nguyen, H. (1977). On Fuzziness and Linguistic Probabilities. *Journal of Mathematical Analysis and Applications, vol 61* , 658-671.
- Puri, M., & Ralescu, D. (1986). Fuzzy Random Variables. *Journal of Mathematical Analysis and Applications vol 114* , 409-422.
- Ross, T. (2004). *Fuzzy logic with Engineering Applications, 2th Edition*. New Mexico, USA: John Wiley & Sons Ltd.
- Runi, J. (2006, Noviembre 4). Fuzzy Time Series Forecasting. *Developing a new forecasting model based on high order fuzzy time series* . Aalborg University Esbjerg.
- Song, Q., & Chissom, B. (1993). Fuzzy time series and its models. *Fuzzy Sets and Systems, vol 54* , 269-277.
- Sullivan, J., & Woodall, W. (1994). A comparison of fuzzy forecasting and Markov modeling. *Fuzzy Sets and Systems, Vol 64* , 279-293.
- Tanaka, H., Okuda, T., & Asai, K. (1976). A formulation of fuzzy decision problems and its application to an investment problem. *Kybernetes, vol. 5* , 25-40.
- Van Dorp, R., & Kotz, S. (2003). Generalized Trapezoidal Distributions. *Metrika Vol. 58* , 1-16.
- Willks, S. (1963). *Mathematical Statistics, 2th Edition*. London: John Wiley & Sons, Inc.
- Zadeh, L. (1965). Fuzzy sets and Systems. *System Theory* , 37.